Using Reproducible Research Pipelines to Help Disentangle Health Effects of Environmental Changes from Social Factors

by

Ivan C. Hanigan

August, 2016

A thesis submitted for the degree of Doctor of Philosophy
of The Australian National University.

Supervisory Panel:

Professor Colin Butler, Chair
Dr Philip Kokic
Dr Rohan Nelson
Dr Steven McEachern
Dr Kathryn Glass

© Copyright by Ivan C. Hanigan 2016
All Rights Reserved

Declaration

I declare that this thesis is my own work, and that I have acknowledged all work of other people. Where I have included the results of collaborative team work, I declare the extent of my contributions in an accompanying statement, to which my co-authors have given signed approvals.

Signed .....

Compiled on 24/08/2016 at 10:36pm

Abstract

The scientific questions motivating this thesis relate to the health effects of environmental changes including droughts, bushfires, woodsmoke, duststorms and heatwaves. Such questions require us to attempt to disentangle health effects of environmental changes from social factors as all diseases have multiple causal factors. Environmental exposures should be explored in the context of many other variables that comprise the biological and socioeconomic milieu. Investigators often narrow the focus to a single environmental cause and health effect. A simple example is bushfire smoke and its direct effects on cardiorespiratory disease. A more complex example is the indirect relationship between drought and suicide. Even simple questions require consideration of numerous putative causes and confounders. Adequately controlling for all these factors in statistical models is difficult. Furthermore results might be sensitive to choice of analysis procedure, or otherwise error-prone due to the many steps.

Such difficulties have led to what some researchers assert is a 'reproducibility crisis' where many scientific publications are difficult or impossible to reproduce. This, with fallacious findings, harms scientific credibility. Reproducibility of data analysis is defined as the ability to recompute results, given a dataset and knowledge of the method's steps. A key problem impairing reproducibility is inadequate documentation of the numerous steps and decisions required for the computations. Reproducible research pipelines allow data and software (such as analysis code) to be disseminated with publications, enhancing reproducibility. However, this approach often places a considerable burden on the analyst. This thesis identifies effective methods for implementing reproducible research pipelines in environmental epidemiology, aiming to reduce this burden.

In addition to the contribution to methodology which this constitutes, the thesis also includes a range of peer reviewed papers (along with accompanying datasets and software packages of code) published by the author, which also add to knowledge. Key findings include health effects of environmental changes relevant to debates

about climate change. Reproducibility of these findings enhances their credibility in response to the heightened scepticism of those debates. Important insights included the finding that the risk of suicide in New South Wales increases in rural men during drought but decreases during droughts for women. Another striking finding was that while bushfire smoke and duststorms each increased cardiorespiratory mortality risk in Sydney, they appear to do so in different ways, with dust having a much higher risk estimate than biomass smoke.

In cases such as this where findings are novel, unexpected or contradict accepted opinion, the scientific method stresses the need for scepticism and critical review. Reproducible research pipelines strengthen our ability to conduct such review beyond what was available in the traditional research model. Not only does the use of pipelines make methodological choices and assumptions more transparent; doing so also safeguards against data misuse by making errors easier to find. Encoding analysis steps in a computer 'scripting' language and distributing the data and code with publications aids readers to assess (and challenge) each choice of data or methods. This will help minimise mistakes in the execution or interpretation of research.

Acknowledgments

In the course of this PhD I was lucky enough to have a multidisciplinary team environment. I thank all my colleagues. I am especially grateful to Prof. Tony McMichael for igniting the spark of my interest in environmental change and human health. Deep thanks go to my supervisor Prof. Colin Butler for fanning the flames and to the National Centre for Epidemiology and Population Health who provided a hearth for the fire. My other supervisors Dr Philip Kokic, Dr Rohan Nelson, Dr Steven McEachern and Dr Kathryn Glass all provided valuable guidance and I thank them. My sincere gratitude goes to Mr Kenneth Johnson, Mr David Dumaresq and Dr Robert Dyball of the Australian National University undergraduate programme in geography and human ecology for providing the substrate of theory and technique I needed to venture into environmental epidemiology. I am also grateful to Prof. Michael Hutchinson, Prof. Keith Dear, Dr Fay Johnston, Prof. Geoff Morgan, Dr Alan Wade and Dr Mark Clements for their assistance at various points. I want to thank all of my co-authors. Being part of multidisciplinary teams is something I find very inspiring. The support of my family Radha, Nea, Maeve, Peter and Robyn Hanigan made it all possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Thesis structure

1.1.1 A compilation of papers and other research outputs

This is a 'thesis by compilation', prepared in accordance with the guidelines of the Australian National University. These guidelines state that the works may be journal papers, data, software, or other kinds of research outputs (Australian National University 2013). In addition, introductory and concluding chapters are included to comply with the following requirement:

> An extended context statement demonstrating the relationship between all aspects of the research is also required as part of the thesis. This will include an introduction to the field of study and the hypothesis or research questions, how these are addressed through the ensuing chapters, and a general account of the theory and methodological components of the research where these components may be distributed across separate papers/chapters. The context statement should be in the order of 10,000 words in length.

The common thread of all the works in this thesis is the aim to disentangle the health effects of environmental changes from social factors. The underlying method used to do this was statistical regression modelling of observational data, and the application of reproducible research methods helped achieve this in a rigorous and transparent way.

The specific research questions pursued in the works of this PhD project are grouped into three clusters of papers and other outputs:

- the effects on mental health of drought
- the effects on cardiorespiratory disease from biomass smoke, dust and heatwaves
- the effects on infectious diseases from local environmental variables.

Chapter one begins with an introduction to the general background of the health effects of the environmental changes motivating the studies. Summaries of the literature suggesting evidence for each of the three clusters is presented. Chapter two presents a discussion of the reproducibility crisis, and the questions posed regarding how to effectively implement the method of reproducible research pipelines in environmental epidemiology. The foundations of the statistical regression modelling framework are also described here. The subsequent three chapters present each of the output clusters of papers, datasets and code (software packages). Each paper, dataset or software package is prefaced by a short introduction explaining how the author contributed to the research. The thesis concludes with discussion of the challenges to implementing reproducible research pipelines in epidemiology and suggests possible future directions for research into a) using the method of pipelines, and b) striving to disentangle the environmental from the social determinants of health.

Table 1.1 summarises the research outputs of this PhD and details are given in appendix 1.

Table 1.1: Summary of research outputs

Publication type	Peer-reviewed	Not peer-reviewed
1. Papers	10	
2. Conference presentation, poster and working paper		1
3. Datasets		4
4. Software packages		3

The PhD project has produced 10 peer reviewed journal papers, one conference presentation (accompanied by a poster and self-published working paper), four datasets and three software packages. All 10 papers were published in high-quality, peer-reviewed journals, two with impact factors higher than nine. The data and software are published without traditional peer-review. Instead these were self-published as open source material on the internet and as such can be reviewed and challenged by any critic wishing to do so. Open access to data and software is a mechanism of scientific review which is increasingly practiced and accepted.

1.1.2 Authorship contribution statements

The PhD was conducted in an interdisciplinary framework. The author was part of several teams during the course of the project, working sometimes as the data analyst with a focus on computing and statistical techniques. At other times he contributed substantive topic expertise. This intersection of computer skill, scientific knowledge and statistics is known as 'data science', and is the key contribution made to all papers incorporated into this PhD thesis. The author is the lead of one paper and second in five others reflecting the substantial contribution made. Each of the papers, datasets and software packages are accompanied by statements that describe the contributions of the author. Approval was obtained and signatures gained from each co-author to signify that they agree to the statement of the author's contribution. Four of the 33 co-authors could not be contacted despite several attempts.

1.1.3 Motivating scientific questions in an eco-social framework

The range of health effects of droughts, bushfires, woodsmoke, duststorms and heat-waves are important and not well understood. They are confounded and influenced by complex inter-relationships with social factors. The great challenge of epidemiology is to make valid conclusions concerning not only the existence but the importance of genuine, causal relationships. The current ability of epidemiology to understand

interlinked environmental and social causes is limited. There is a need to develop new methods that utilise an eco-social framework (McMichael *et al.* 2015). The work reported in this thesis addressed this challenge by applying data analysis methods that disentangle health effects of environmental changes from social factors. The following three sections will provide background literature that places the health risks of each of the three clusters of health outcomes and environmental exposures into their historical context and describes the research questions that were investigated by the papers.

Before the health-effects literature is introduced, a brief discussion is required of the conceptual model of causality that is used in an eco-social framework. The aim here is not to provide a comprehensive review of the literature on causal inference (which extends well beyond epidemiology into philosophy, metaphysics and logic) but only to describe highlights of the conceptual model in general terms. At the foundation of the approach used is the understanding that causation is an interpretation, not an entity. As explained by Lucas & McMichael (2005):

> we can never know absolutely that exposure X causes disease Y. There is no final proof of causation: it is merely an inference based on an observed conjunction of two variables (exposure and health status) in time and space.

There is also the important realisation that an holistic analysis is needed, one which accounts for a multitude of relationships of environmental and social indicators with health and wellbeing outcomes. This is because disentangling complex interrelationships requires us to investigate complex, sometimes protracted, causal pathways, confounding variables and effect modifiers both proximal (causes close to the effect) and distal (causal factors further upstream on the causal pathway) (McMichael 1999).

This is sometimes conceptualised as a 'web of causation', and causal pathways conceived as chains of linked causes and intermediary effects, culminating in the health outcome of interest. This metaphor of the 'web' has been criticised as limited (and constraining), and an alternative 'ecosocial' model has been proposed (Krieger

1994). However, whichever conceptual framing dominates epidemiological theorising, there is a tendency for epidemiologists in practice to reduce the focus of their study to deal with simple datasets, only a few variables, and intuitively plausible exposure-response relationships. McMichael (2013) argued that when the focus of study turns to health effects of environmental change the traditional mode of research risks 'epidemiologising, (and thus truncating) the topic'.

Instead, he proposes that epidemiological researchers:

> spread their wings and learn about new concepts, methods, the dynamics of genuine interdisciplinary research, and how to handle and communicate unusual levels of uncertainty.

Such an approach however will likely lead to substantially increased numbers and diverse ranges of inputs to the data analysis. As can be seen from the literature summarised below, the numerous putative causal mechanisms and potential data inputs to be considered are multiplied substantially compared with more traditional epidemiological studies. Methods that allow investigation of 'multiple working hypotheses' (Chamberlin 1890) and employ 'strong inference' (Platt 1964) are needed. These will enhance our ability for 'explicitly considering competing hypotheses, and if possible working to develop experiments that can distinguish between them' (Nuzzo 2015). Unfortunately, experiments are difficult or impossible in most epidemiological research settings, which are usually in inherently 'noisy' environments of dynamic human populations and ecosystems.

In response, beyond experiments, a range of other techniques and considerations for design and interpretation of studies are available to support causal inferences (Bradford Hill 1965). It should be noted that statistical models alone, including those applied in the papers of this thesis compilation, can never fully disentangle distinct health effects of environmental and social changes.

The approaches taken in this thesis to address the limitation of statistical models as a tool to determine causality are consistent with recommended statistical practices. Gelman & Hill (2007, p.171) write:

> Estimating causal effects requires one or some combination of the following: close substitutes for the potential outcomes [counterfactuals], randomization, or statistical adjustment.

In addition, Gelman & Hill's recommendations to use Bayesian and frequentist tools to help refine causal inference using regression and multilevel models are described in Chapters 9, 10 and 23 of their textbook (Gelman & Hill 2007). However, even the application of such methods does not resolve all limitations. In particular the papers compiled in this thesis were prone to the 'problem of multiple comparisons', increasing the risk of false positives. These limitations can be lowered by increasing the number and quality of independent investigations that 'address a hypothesis and build evidence for or against it' (Peng, 2011). An example is presented in paper 8 of this thesis, in which our study of a reduction in death rates associated with reduced air pollution in Launceston, Australia, replicated similar findings from reduced smoke pollution in winter from biomass in a small community in Montana, USA, the temporary closure of a steel mill in Utah, USA, the reduction of the sulphur content of petrol in Hong Kong, China, and the cessation of coal sales in Dublin, Ireland. In turn this type of replication study will be aided by the use of reproducible research pipelines, as well as by enhancements to the study design, moving toward randomisation and experiments where possible. Indeed, the value of this approach is a key contribution to knowledge made by this thesis.

Other recent methodological advances include the work of Pearl et al. (2016) who proposed, in addition to conventional statistical methods, what they call '*extra*-statistical' methods. These are defined as techniques that 'formally articulate causal assumptions—that is, to create causal models . . . [and] . . . link the structure of a causal model to features of data' (p. 5). In the future it may be possible to use such probabilistic and causal reasoning to estimate causal inferences in ways that partly

or even fully overcome the inability to compare counterfactual potential outcomes in observational data.

In conclusion, the eco-social framework pushes through boundaries around the scope of epidemiological studies that have been rigid for a long time. It relies on the strong foundations of such heuristics for causal reasoning as those outlined above to ensure that any novel and unexpected findings from these increasingly complicated studies are assessed properly; to avoid risks of publishing more fallacious associations mistaken as causal relationships.

1.1.4 Literature supporting the three clusters of research questions

This section highlights parts of the literature available supporting each of the three clusters of research questions addressed by the thesis. The aim is not to provide a comprehensive review of the literature for each issue because the published papers provide their own background material to contextualise the particular epidemiological questions. Instead, each of the clusters of health effects and environmental exposures are first characterised in terms of the main health outcomes and exposure pathways, the multiple interacting environmental and social factors, and the magnitude of public health impacts related to the issue. The epidemiological evidence of health effects related to the exposures of interest are summarised in general terms. Specific exposed populations and sensitive subgroups are discussed while uncertainties or limitations of the literature are noted.

1.1.5 The effects on mental health of drought

Chapter 3 of this thesis is composed of a cluster of research related to the effect of drought on mental health problems. Drought is defined here as a prolonged period of abnormal dryness in an ecosystem. There are a number of definitions and indices of drought, which generally reflect different perspectives. A useful classification is by Holton et al. (2003). They identified four main types of drought. Climatic drought occurs when there is a prolonged period with below average precipitation. Agricultural drought is when there is insufficient moisture available for average agricultural production. This can happen even with average precipitation, due to soil conditions or agricultural techniques. Hydrologic drought occurs when the water reserves available in sources such as aquifers, lakes and reservoirs fall below the statistical average. This also can happen with average precipitation when increased usage of water diminishes the reserves. Finally, there is socio-economic drought occurring when the overall societal demand for water exceeds the supply.

Several drought indices exist. When considering them in relation to suicide it is

important to link the impact of the drought with that of human livelihoods in the area studied. Smith et al. (1992) contend that climatic and agricultural droughts are the most relevant to impacts on human livelihoods. An additional complexity is that droughts are often accompanied by high temperatures (Mazdiyasni & AghaKouchak 2015). Chronic stress or acute shocks from heatwaves or (in some areas) bushfires may push individuals or communities toward, or even beyond, thresholds that social and ecological systems can cope with or adjust to.

In recent decades there has been substantial concern within Australia that drought impairs the mental health of rural populations. Mental health outcomes of interest include anxiety, depression, distress and suicide. The topic has frequently been raised by the media, by rural politicians and by mental health support groups (Australian Broadcasting Commission News 2006).

Despite this concern, however, the number of studies that have examined the putative relationship between mental health in general (Stanke *et al.* 2013) or suicide specifically is surprisingly poor (Vins *et al.* 2015). On the other hand many papers have explored links between suicide and other (non-drought) climate variables, such as temperature. Two major reviews analyse this literature (Deisenhammer 2003; Dixon & Kalkstein 2009). Although, as mentioned, studies of drought and suicide have been rare, there are studies of some climatic variables related to dryness. Preti (1998) reported higher suicide rates in drier towns in Italy. However, the findings of these studies reveal confusing, sometimes contradictory associations between climatic variables and suicide (Kim *et al.* 2011).

Several exposure pathways have been hypothesised by which unusually dry conditions may worsen mental health problems (Vins *et al.* 2015). The best accepted is that the most potent causal pathways operate through economic downturns, sometimes with out-migration effects. However, it is also plausible that close witness of a suffering and even dying landscape, including wildlife, vegetation, livestock, and eventually even home gardens affects cultural identity. People with a strong sense of connection to the land or who are drawn to environmental stewardship may be particularly vulnerable.

The direct economic effects occur because droughts increase the financial stress on farmers and farming communities (even if partially compensated by drought relief, interest relief and welfare payments). Such difficulty may be in conjunction with other economic stresses, such as rising interest rates, falling commodity prices, or an unfavourable foreign exchange rate. Reduced farm incomes then depress economic activity in rural towns. In some regions entire economies may be affected, accelerating migration to metropolitan areas; weakening and stressing social support systems and lessening social interaction. In some cases rural depopulation may pass a tipping point of a different kind, leading to a self-reinforcing loss of critical services, such as hospitals, schools and doctors.

Other causal pathways include the psychological toll following environmental degradation (Speldewinde *et al.* 2009). Human experience of drought may take into account aesthetic values, recreational pursuits and the effects of uncertainty or variability. Concepts such as 'solastalgia' (Albrecht *et al.* 2007) which captures climate and other related feelings of uncertainty and loss, may be relevant. Solastalgia has been defined as a lack of solace in one's home environment caused by the impact of negative environmental change. This may be acute during droughts linked with decisions to sell or kill starving animals or to destroy orchards and vineyards, which in some cases have been painstakingly accumulated over generations. Such loss, and even the apprehension of loss, seems likely to further burden the mental health of farmers and their families. Furthermore, such stress may extend to other sections of the community also affected by long-term environmental degradation. The experience of seeing suffering wild plants and animals, or parched urban parks and gardens, and contemplation of their loss is likely to be extremely painful for some individuals, such as those with high biophilia (Wilson 1984), for example among carers of injured wildlife.

As mentioned few studies have examined the relationship between suicide and drought. One analysis of annual suicide rates in the Australian state of New South Wales (NSW) found an association between suicide and year-to-year decline in annual rainfall between 1964 and 2001 (Nicholls *et al.* 2006). In that study, to

which I contributed prior to enrolling in my PhD, a decrease of 300mm of rain was associated with an increase in suicide rate of about 8% above the mean annual rate. Another study of NSW, for the period 1901-1998, focussing on the association between conservative government and suicide also found an association between suicide and drought (Page *et al.* 2002). The authors argued that conservative government programmes (or the perceived prospects farmers might expect under a conservative government) might influence mental health directly, or that a correlated increase in anomie (decreased inclusiveness of society) and the associated lowering of social capital might enhance the risk of suicide in vulnerable individuals. The authors controlled for drought (among other things), and found that years in drought were associated with an increased suicide risk of about 7% for men and 15% for women, across the whole population. In opposition to these results, Guiney (2012) reported finding no pattern of increasing farming suicides during periods of drought. However, the context of that result, and the value of this study for causal inference is challenged by methodological flaws, including its design and sampling procedure. In particular, the study did not include a non-drought period for comparison.

In contrast, paper 1 of this PhD (Hanigan *et al.* 2012) did find an association between drought and suicide. The paper reports an analysis of 38 years of suicide data (1970-2007) linked with a new drought dataset created specifically for this research (see dataset 1 and software package 1 of the thesis). The focus of the study was on the effects of drought on farmers and farm workers. The drought exposures were calculated from climatic data for 11 subregions of NSW, and stratified by rural/urban region, age and sex. A strong association was observed in rural males, particularly for the 10-49 age group. However, in contrast to our original hypothesis the suicide risk for rural females aged over 30 was found to decrease during droughts. This is in contrast to the findings of Page et al. (2002), but is probably explained by the fact their study period included the 'sedative epidemic' from 1960-1967 (Oliver & Hetzel 1973) a period of easy availability of barbiturates, during which female suicide rates were much higher than they were subsequently. It is speculated that females, especially young women, overdosed on barbiturates. This may have been

often without any deep intention to die. Many drugs are lethal in overdose, but barbiturates appear to have been particularly risky. In any case, much greater regulation of this class of pharmaceuticals appears to have greatly reduced the female suicide rate, in both urban and rural settings, and not only during droughts. It is likely this reduction in accessibility to sedatives may have reduced both the depressed person's potential overdoses, and the suicidal person's access to the lethal means they would need for suicide. This may have reduced the risk substantially in women in general, and especially during droughts in the period of the study from 1970-2008.

Further support for the hypothesised effects of drought on mental health is provided by the results from paper 2 of this thesis (O'Brien *et al.* 2014) in which a distress indicator from a survey was investigated. That study explored various characterisations of drought and found that distress levels were higher during long and constant drought for rural people, but not for urban residents.

Another investigation of specific exposed populations and sensitive subgroups have revealed complicated patterns of drought effects, potentially mediated by social and other non-environmental factors. For example Powers et al. (2012) found that adverse climate events (including droughts, floods or fires) did not appear to affect Australian women's wellbeing. That result was further corroborated by paper 3 in this thesis (Powers *et al.* 2015) which found a lack of association between drought and mental health in Australian women. There may be a complicated causal pathway involving gender which modifies the effects of drought identified in these studies.

1.1.6 The effects on cardiorespiratory disease of bushfire smoke, woodsmoke, duststorms and heatwaves

Chapter 4 of this thesis comprises a cluster of research related to the cardiorespiratory health effects of bushfire smoke, woodsmoke, dust and heatwaves. Particles (and perhaps noxious gases) in smoke and dust directly influence the respiratory system through their inhalation and lodgement in the lungs. Particles may then affect the cardiovascular system after their entry into the circulatory system from the alveolae.

One paper in this cluster investigates heatwaves which, while not in the same class of exposures as particles and gases, has causal influences that are closely related to those of the pollutant compositional attributes of the atmospheric environment.

For decades, researchers have studied the public health impacts of ambient outdoor air pollution, particularly from the effects of particulate and gaseous pollutants, especially associated with the combustion of coal, petroleum and biomass used for cooking (Pope & Dockery 2006). Far fewer studies have examined the effect of intermittent woodsmoke such as that which occurs in bushfires, or from woodsmoke trapped by inversion layers during winter months as wood is burned for heating (Smith 1993; Naeher *et al.* 2007).

Epidemiological literature that investigates the relationship between health and air pollution exposures has primarily used time-series methods that study variations of some health outcomes such as deaths or hospitalisations from specific disease groups (Peng & Dominici 2008). These outcomes are usually monitored by day across whole cities, and relationships with atmospheric variables estimated in regression models. These typically focus on ambient air pollution measured by a network of monitoring sites scattered across a city, time matched to the health outcomes on the same day or a few days after. These models often also consider interactions between air pollution, temperature and humidity, and control for other putative confounders such as age, gender, day-of-week, public holidays and influenza epidemics (Peng & Dominici 2008).

The air pollution literature reports hundreds of published time series studies. Considerable research effort has been spent searching and testing statistical methods which most adequately adjust for confounders that vary over time. In 2002 a methodological flaw in the use of a common computer algorithm was uncovered (Dominici *et al.* 2002; Lumley & Sheppard 2003) which cast doubt on many earlier papers. This led to a reanalysis of results originally reported in 37 published original articles and reports (Health Effects Institute 2003). This reanalysis still supported the previous generally accepted conclusion that particulate matter in air harms some health outcomes. This evolution of the literature has generated an increasingly robust

evidence base linking cardiorespiratory disease to air pollution (especially fossil fuel induced pollution from emissions from factories, coal burning power stations or transport).

However, a gap in epidemiological knowledge remains concerning the health effects specifically attributable to acute exposure to biomass smoke from bushfire events and from spikes in smoke from wood-burning heaters worsened by smoke-trapping inversion layers. There is also a dearth of literature investigating acute exposure to crustal material blown over cities during duststorms. This research field has been hindered by the unavailability of adequate datasets that explicitly pertain to these sources. This PhD incorporates a new, open and extensible database developed by the author in association with several collaborators to identify historical spikes in particulate matter concentrations and to evaluate whether they were caused by vegetation fire smoke or by other means. The protocol for developing this database is published as paper 4 of this thesis (Johnston *et al.* 2011b). The database itself is included as a research output (dataset 3) and distributed in an open online format that allows the research community to extend the history of these events. Software that was developed to create that database is published as open source code in software package 2.

The subsequent research on the health impacts of these air pollution events that the database enabled is included as papers 5 to 8. Paper 5 linked mortality to bushfires and duststorms, finding evidence that this exposure harms cardiorespiratory health (Johnston *et al.* 2011a). Paper 6 investigated the relationship between morbidity (hospital admissions and presentations) and bushfires, again finding evidence of negative health effects (Martin *et al.* 2013). Paper 7 is slightly different. It explores the relationship between heatwaves and a wider range of mortality and morbidity outcomes (Wilson *et al.* 2013). This paper concluded that the largest public health impacts associated with heatwaves are increased cardiorespiratory diseases. Paper 8 reports the results of an investigation of an intervention that reduced air pollution in the Australian city of Launceston in the state of Tasmania (Johnston *et al.* 2013). That intervention was associated with a substantial reduction in mortality and

supports the view that governments can improve population health outcomes if they act effectively to reduce woodsmoke pollution.

1.1.7 The effects on infectious diseases of local environmental factors

The research included in Chapter 5 of the thesis investigates environmental determinants of some infectious diseases, including bacterial gastroenteritis and Ross River virus. There is a high level of public anxiety in Australia related to these conditions, and their putative links with environmental changes including to climate, weather and vegetation.

Paper 9 of this thesis (Hall *et al.* 2011) investigated the link between local weather patterns and infection by gastroenteric pathogens, controlling for social factors such as age, income and health insurance status. These pathogens can cause diarrhoea, vomiting, dehydration and death. Though mortality in Australia from gastroenteritis is uncommon it still causes a considerable burden in terms of lost productivity and cost.

Paper 10 examined local variables such as the aquatic environment and proximity to waterways, hypothesised to determine habitat quality for the mosquito vectors of Ross River virus (Vally *et al.* 2012). This vector-borne disease causes few deaths but has significant impacts on wellbeing and the economy, due to lost productivity. In some people it can generate arthropathy persisting for months (Harley *et al.* 2001). The study sought to disentangle the different environmental exposure pathways, regulated by social factors, that influenced the physical environment and human behaviour patterns. These may have led to differences in exposures across the local region. This topic is of public health relevance because in some parts of Australia there are continuing outbreaks of Ross River virus, and new housing developments are sometimes located in high risk areas close to waterways and coastal wetlands.

The epidemiological evidence that the incidence of these diseases is related to the exposure via local environmental changes cannot be simply summarised.

Complicating this aspiration is that there appear to be multiple interacting hazards related to the complex interplay of environmental variables which determine the suitable habitat for pathogens (and vectors), and also the human exposures to these (Pavlovskii 1966). Social variables which mediate these exposures (and influence the occurrence of the environmental habitat factors) include income levels and remoteness. For instance, such populations may have poor community facilities including sewerage and water supply. Many other socioeconomic and demographic factors exist which may modify effects or confound the association, if any, between environmental variables and infectious disease.

The recognition that the incidence of many infectious diseases is seasonal suggests the importance of environmental factors. There are also regional differences in timing and magnitude of infection peaks. It has been suggested that local weather conditions can potentially influence human infection risk by affecting local environmental pathogen proliferation and carriage. Such carriage or multiplication might occur in locally produced or contaminated food or water, or via animals or elsewhere in the general environs. Weather can also affect human behaviour, including food preparation and diet. Further information is presented in paper 9 of this thesis (Hall *et al.* 2011).

Environmental epidemiological evidence linking health and Ross River virus includes the relationship between epidemics preceded by high tides or unusually heavy rain in late spring and summer, enabling surging populations of the vector *Aedes* mosquitoes in seasonal wetlands and tidal salt marshes. Additional discussion is provided in paper 10 (Vally *et al.* 2012).

The papers included in this thesis addressed these diverse questions through the use of high quality surveillance data, quantifying the magnitude of the health outcomes in sensitive subgroups. Environmental data were collected that represented the potential key exposures, with both spatial and temporal attributes taken into account. This allowed statistical regression modelling to be conducted that estimated associations while controlling for different environmental and social factors.

1.1.8 Conclusions regarding the epidemiological literature

A traditional aspiration of epidemiologists has been to disentangle the roles of the different causal agents. However, the focus on isolating causal factors for complex phenomena is difficult and sometimes too reductionist. A raft of causal agents may interact in a variety of ways and it is important to always be cognisant of that (Krieger 1994). A general definition of environmental epidemiology is that it attempts to explain the effect on human health of environmental phenomena amongst a complex web of inter-related non-environmental factors. The 'environment' is obviously a broad concept, and can be interpreted as having social as well as inert physical and infectious components. In this thesis, the focus is on human responses to stimuli from their physical surroundings, such as the air, water, ecosystems and geology (Meade & Earickson 2000; Millennium Assessment 2003). More specifically, studies in environmental epidemiology often attempt to narrow the focus to a single environmental cause and its concomitant health effect. These reductionist questions are usually explored in the context of many other known or suspected causes such as human biological parameters and the socio-economic milieu. Ideally, epidemiologists strive for study designs that will allow them to identify the most important causal agents; not only pertaining to the environmental cause under investigation, but also from factors attributable to human biology and to societies.

Chapter 2

Reproducible research

2.1 The reproducibility crisis

Reproducibility is defined as 'the ability to recompute data analytic results given an observed dataset and knowledge of the data analysis pipeline' (Leek & Peng 2015). This definition distinguishes reproducibility from replicability which is 'the chance that an independent experiment targeting the same scientific question will produce a consistent result' (Leek & Peng 2015).

Replication refers to the case where a study is conducted in the most similar conditions as a prior study to investigate the 'consistency' of the effect (*sensu* Bradford Hill (1965)). This means that, according to Bradford Hill, we might have more confidence in a causal explanation for an association if the same answer had been achieved in a variety of different situations. Reproducibility on the other hand is the ability to easily recalculate a result from any prior data analysis that is repeated on the same data and using the same analytical steps. This simple sounding requirement of re-computing something that has already been done once, is deceptively complicated in practice. Many of the computations of the type used in the data analyses being discussed are in fact very difficult to reproduce.

Some researchers assert that much scientific research is difficult or impossible to reproduce, and that this harms scientific credibility (King 1995; Peng 2011). This

has been termed the 'reproducibility problem' (Buck 2015) or even a 'crisis' (Peng 2015). Various solutions have been proposed for this problem, such as developing new professional guidelines and standards about the transparency of methods (Nosek *et al.* 2015), changing academic career incentives (which currently have a perverse sense of encouraging obfuscation by authors so that they can retain competitive advantage) (Alberts *et al.* 2015), changes to journal requirements to enforce stricter rules regarding supporting information of the claims made in submitted manuscripts (Peng & Eckel 2009; Irving 2015; Loder & Groves 2015), and even a move to pre-register a data analysis paper, where journals can review and accept papers prior to data collection even having occurred (Nosek & Lakens 2014). The rationale for this is based on the assumption that if the protocol is then adhered to, the study will be more likely to be reproducible because the incentives to adjust or manipulate data analysis procedures are removed.

These approaches are known in general as 'reproducible research', and includes developments such as software tools that keep a log of all work conducted through a data analysis pipeline.

2.2 Reproducible research pipelines

It is asserted in this thesis that the method of using reproducible pipelines addresses the reproducibility crisis and helps to improve scientific credibility. This is achieved using rigorous documentation of steps in the analysis, together with making this information available with publications (including the data and software code used in each study).

One of the most active and fruitful areas of research in this field has been on computational methods that allow scientists to more easily log all the steps they make while conducting data analyses. Software engineering solutions have been developed that combine both the data analysis code and the descriptive prose that constitutes the publishable report into a compendium (Gentleman & Temple Lang 2004). Compendia have been used to produce journal publications (Schulte *et al.*

2012), journal supporting information documents - such as that included with paper 1 of this thesis (Hanigan *et al.* 2012) - and even books (Peng & Dominici 2008). But it has also been recognised that reproducible research can still be wrong and a prevention approach has been recommended that incorporates evidence-based data analysis tools and techniques into such a pipeline (Leek & Peng 2015).

2.3 Pipelines help documenting data analyses

Adequately documenting the methods and results of data analysis helps safeguard against errors of misexecution and misinterpretation. This thesis proposes that reproducible research pipelines address the problem of adequate documentation of data analysis.

The core concepts and flow of steps in the method are shown in Figure 2.1 (after Peng et al. (2006) and Sólymos & Fehér (2008)). In this model there are two main actors: the author and the reader. The author moves from left to right, from initial hypothesis and study design, through data collection and pre-processing, to analysis and reporting. The aim is to conduct all steps of the analysis work in such a way that the key dataset and code script can be sent through the distribution mechanism and the reader can easily move from right to left. Thus the reader can start with the published results and then dig deeper by assessing the analysis code and analytic data to gain full understanding of the methods steps.

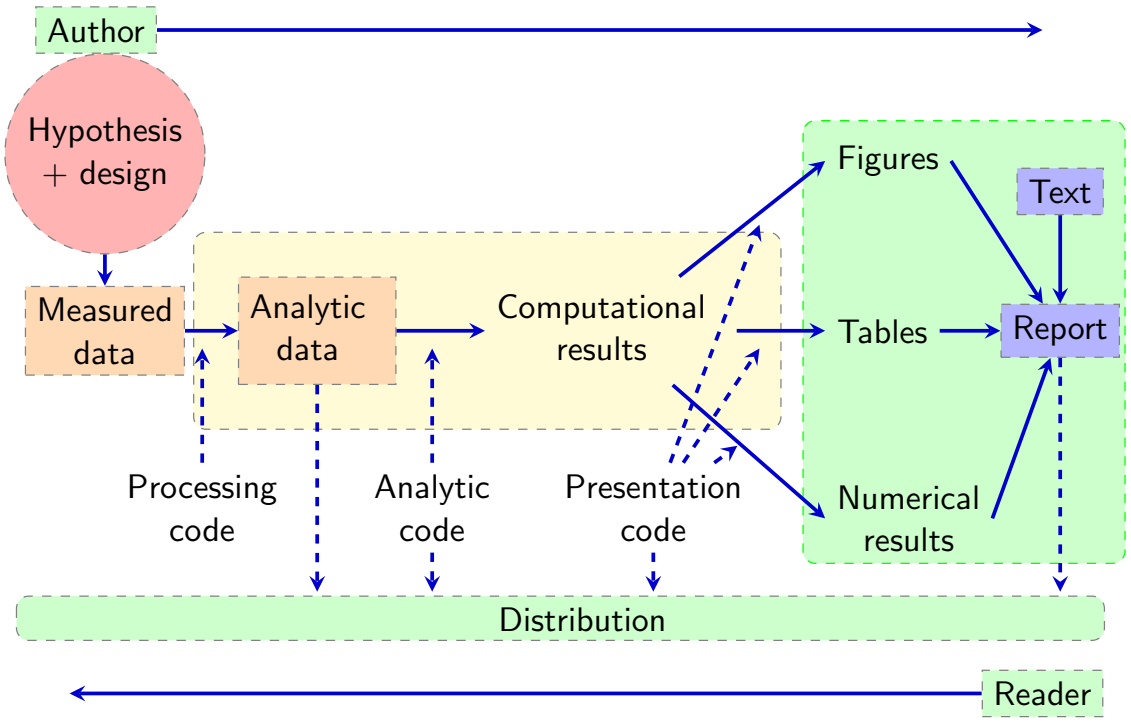


Figure 2.1: A schematic flow chart showing a pipeline to prepare a reproducible research report.

As noted above reproducibility is not the same as replication because the former relates to a specific data collection and analysis process, whereas the latter requires new data to be collected. These concepts are closely linked though and the connection between the two is shown in Figures 2.2 and 2.3. These highlight that the minimum requirement for reproducibility is for the specific dataset and analysis code used for a report be available, while replication is made even easier if data collection and analysis can use the reproducible dataset and code as substrate to build the new work on.

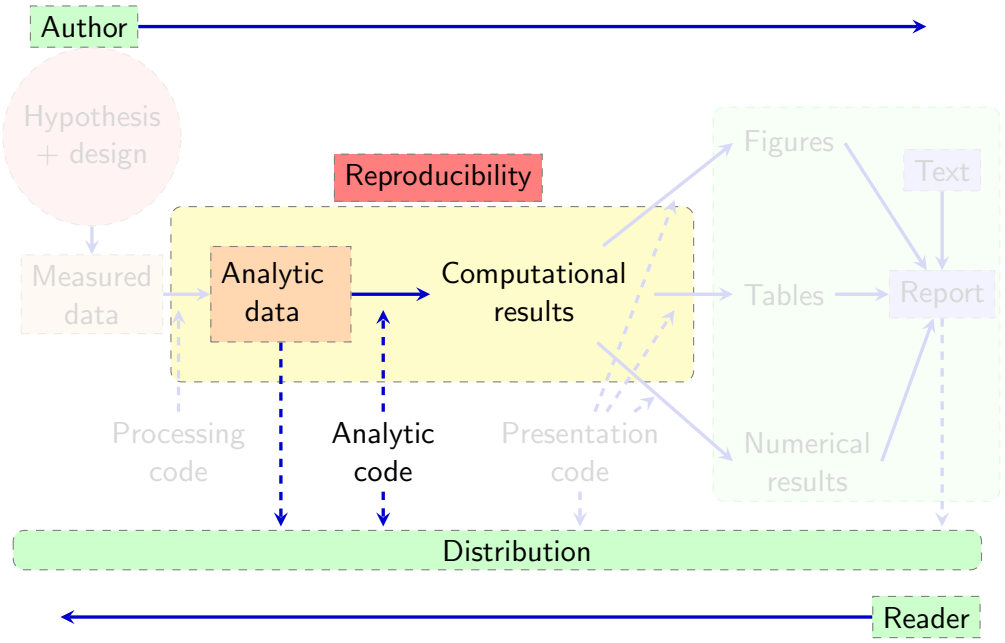


Figure 2.2: A schematic showing the minimum requirement of reproducibility

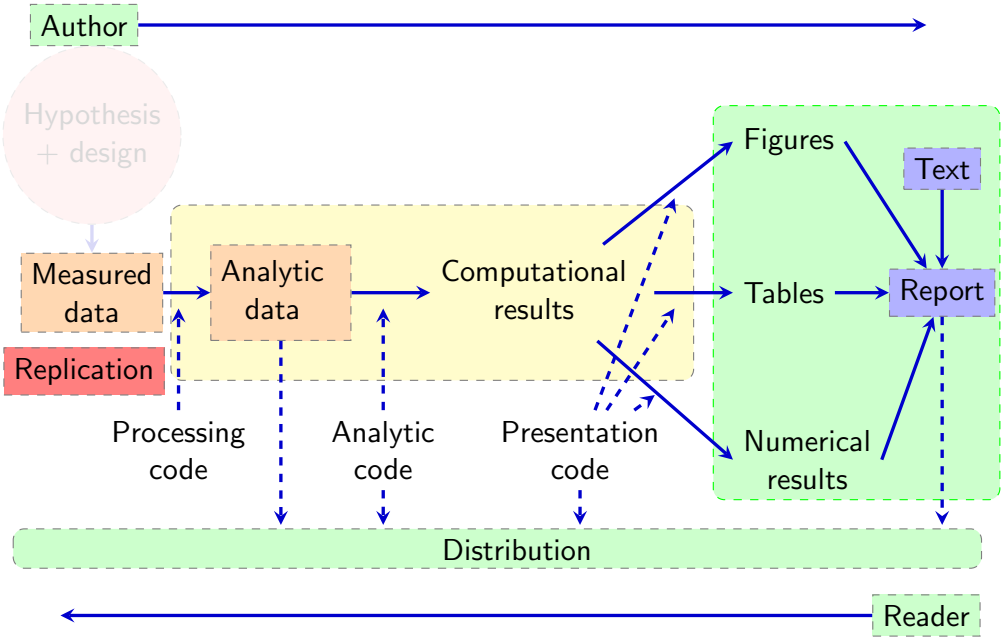


Figure 2.3: A schematic showing the relevance of pipelines for replication

Five governing principles of the method of reproducible research pipelines (derived from Peng et al. (2006)) are stated here and expanded on in appendix 2. These are:

1. Hypothesis and design.
2. Data:
    - Measured data (collected, cleaned and checked)
    - Analytic data (pre-processed).
3. Analysis methods:
    - Computational results (for figures, tables and numerical results)
    - Coding for data processing, management and analysis of all associated data and reports.
4. Documentation:
    - Reports (text and formatted manuscript)
    - Metadata.
5. Distribution of data and code for archiving, or access by others.

Reproducibility helps safeguard against associations being mistaken as causal when they are not. This is because the method of pipelines improves the transparency and rigour of data analysis and makes it easy to audit data analyses and make quality control checks. Assumptions are easy to challenge and results verified by new analyses. Reproducible research pipelines extend traditional research. Researchers who use pipelines encode all their steps in a computer 'scripting' language and distribute the data and code with publications.

This is needed to provide the trust that a data analysis was appropriately conducted and avoided errors of misexecution, or methodological design. The types of error that can come from the execution of data analysis include failure to spot typographical errors in code and data, whereas errors that are more methodological include the selection of an inappropriate method used for a given data type (for

example fitting Poisson models to non-count data such as mortality rate ratios). Reproducibility, if accompanied by transparency and critical review is an important tool for screening and rejection of poor data analyses by referees, editors, and other scientists in the community.

It is often difficult to produce reproducible research due to multiple revisions on datasets and analysis plans. In essence the contribution of pipelines is toward a set of tools that enable analysts to deal with the issues of tracking the many choices through a complex path of stepping stones between data, analysis, reporting and archiving activities. A technical discussion of effective data management procedures for reproducible research pipelines is presented in appendix 2.

2.4 Problems arising when studies do not use pipelines

There have been notable examples of studies that have found it difficult to reproduce published results, and also some notable examples with results that were actually errors. Four examples have been selected for discussion in this section, ranging from trivial (and potentially inconsequential) to serious errors with grave implications for government policy and expenditure of public money.

In the first example from the world of bioinformatics (Xie *et al.* 2011) a rather trivial problem is evident in which a published abstract includes the statement: 'we have used (insert statistical method here) to...'. This is no doubt an embarrassing error for the authors but is really only a minor technical error caused by a disconnection between the author writing the report and the author conducting the statistical analysis. In a reproducible pipeline such an error would be hard to make because the analysis method is closely connected to the report of the results.

The second example is more worrying and comes from the domain of chemistry. Many pages into the supporting information document of a paper by Drinkel et al. (2014) there is the statement 'please insert NMR data here! where are they? and for

this compound, just make up an elemental analysis' (page 12). This problem shows the potential for analysts to fraudulently claim that data or analysis results exist when they actually don't. It would be hard to make such an error (or engage in any deceit) in a pipeline because the data and results are tightly connected to the report that is created to describe them.

The third and fourth examples come from the field of epidemiology where an influential paper published by Clancy et al. (2002) reported a decrease of 10% in cardiovascular disease mortality associated with an intervention to reduce the air pollution in Dublin, Ireland. The first problem was due to a confusing statement in the methods section in which the authors state they used 'age-standardised Dublin death rates... ...assuming a Poisson distribution' without explaining that because Poisson regression is applicable only to count data and cannot be applied to rate data, the model needs to be parametrised with the numerator (counts) on the left hand side of the equation and the denominator on the right as an offset. It was also not clear in the description of the method if the age-standardisation method employed the direct or indirect age adjustment method, a choice which can produce large differences in the estimated rates and their comparability with rates in other populations with different age structure. This confusion is due to lack of clarity in the methods section of the journal paper and may have been more easily avoided had the code and data been distributed along with the paper.

The fourth example also comes from this same paper by Clancy et al. (2002) and demonstrates a bigger issue. This was due to the ambiguity in a statement about how the authors: 'controlled for secular changes by adjustment of directly age-standardised rates for the rest of Ireland, excluding Dublin'. During sensitivity analysis post publication the original authors found that the results were very sensitive to the way that reference population rate was constructed. In fact after correcting their methods they found no evidence of a change in cardiovascular mortality; the original estimate of a statistically significant 10% decrease was revised to an estimate of no effect (0.1%, P = 0.98). The key methodological error was the manner in which the adjustment was made for secular trends, with the original using daily time-matched

rates whereas the revision used Loess smoothed reference mortality rates (Dockery *et al.* 2013). Such a methodological issue is not easy to avoid, even when using the method of pipelines. However, it is likely that had the data and code been available for other researchers to test different configurations the issue may have been resolved sooner than it was.

2.5 Numerous causal factors and confounders in protracted causal pathways

It is a truism that diseases have a multifactorial pathogenesis, with multiple causal factors influencing health outcomes. Environmental epidemiology is particularly prone to a large number of potential causal mechanisms that interact. This makes reproducibility difficult, but it also means that well documented analysis procedures are even more necessary to track the permutations and different configurations or sensitivity analyses that might be conducted while sifting through the many variables to be assessed.

An often employed study design is that of observational data analysis (given that intervention studies are usually not feasible), and that is the type of analysis method employed in the studies included in this thesis. This section provides a discussion of some statistical modelling methods that can be used to disentangle the influence of environmental changes from social factors in such studies. It is important to employ reproducible research methods so that future data analysis (utilising the methods of pipelines as proposed) might build on these results. In this way there will be improvements in testing, validating and ultimately understanding the complex patterns that can be found when studying these factors.

A well-known caveat about observational data analyses is that they cannot, of themselves, establish cause-and-effect relationships. This is because associations between variables can arise from dependence between exposures and outcome that may be mediated by hundreds of unmeasured factors. Results can only indicate

the direction and magnitude of the extent that variables are associated with each other, and the ability of the statistical model to explain the variance in the observed data. Inferences and interpretations about a cause-and-effect relationship can only be based on the judgement of the analyst with respect to other pertinent information such as consistency with other studies and theoretical understandings.

The approaches that environmental epidemiologists have developed to investigate the complex interacting components of environmental and social systems can be applied in observational data analysis and can help to provide evidentiary support for putative causal relationships. The epidemiological approaches to causal inference build on the foundations laid by Bradford Hill including his seminal paper in which he proposed nine different criteria to assess causality (Bradford Hill 1965). These nine criteria are:

1. ***Strength***: either as the magnitude of an effect (such as a ten-fold increase in relative risk) or the strength of the study design and power
2. ***Consistency***: the ability to replicate the findings
3. ***Specificity***: while diseases may have more than one cause, if an association is limited to specific groups then the case for a causal association is strengthened
4. ***Temporality***: cause precedes effect, and can be observed in sequence
5. ***Biological gradient***: such as a dose–response curve
6. ***Plausibility***: based on the knowledge-base available, such as the existing literature
7. ***Coherence***: the cause-and-effect interpretation of an association should fit with the known facts
8. ***Experiment***: the results of studies that ask if actions taken on the basis of a proposed cause-and-effect association alter the frequency of the outcome
9. ***Analogy***: for example, by analogy to the known risk of lung cancer in active smokers, persons exposed to second-hand smoke plausibly have an increased lung cancer risk mediated by the same biological pathways.

These criteria have been used widely in environmental epidemiology. They have also been criticised sometimes as inappropriate in practical applications, especially where causes might be subtle and complex systems confound the associations, or in cases where little is known about the system to be able to provide support for some criteria (Lucas & McMichael 2005). Even so, it is notable that Bradford Hill himself did not prescribe these criteria as rules that must be fulfilled before an association can be judged as causal, but as ways of examining if cause-and-effect is the reasonable inference. In recent years it has been common for statistical data analyses to rely on the strength of a 'null hypothesis significance test' p-value, to the exclusion of all other casual criteria, or conversely for the results of observational data analysis to be dismissed as irrelevant because it does not meet the eighth criterion of 'experiment'. This is regrettable because the broad acceptance of the eco-social framework described above will necessarily include studies that span this breadth of the available approaches to causal inference as proposed by Bradford Hill and others.

McMichael (2013) explained that many risks from environmental change are mediated by complex, protracted pathways and there is a risk of 'epidemiologising' the research agenda into simplistic, narrowly focussed, 'me-too' studies. He cautioned against choosing such 'lamp-post' research priorities (a reference to a joke about a drunk looking for their car keys which they lost in the dark, but only searched under the lamp post 'because that is where the light is'). In the research included in this PhD thesis my co-authors and I undertook to understand the effects on suicide from drought. We suggested that future research should consider the causation of suicide using a holistic model known as the 'five-capitals framework' that involves financial, physical, social and human factors together with natural influences, such as climate (McMichael & Butler 2002; Nelson *et al.* 2005).

Suicide is a complex phenomenon with many causal influences (Vins *et al.* 2015). Such an holistic framework as the five-capitals would allow us to systematically investigate how combinations and interactions among various factors influence suicide risk. This conceptualisation includes a more comprehensive range of explanatory variables than we have modelled in the study reported in paper 1 of this thesis. For

example, we included a seasonal term to capture the well-known peak in suicide during spring. A five-capitals perspective might enhance our ability to include variables related to drought that might explain some of this seasonal pattern, such as daylight hours or pollen. The association between suicide frequency and amount of bright sunlight was studied in Victoria, Australia during 1990–1999 by Lambert et al. (2003). This particular variable is plausibly related to droughts due to the lessened cloud cover, and may be linked to suicide rates via mechanisms including the immune system function and related psychiatric effects of vitamin D and serotonin. Additionally there is some literature pointing to the potential causative role of pollen and allergy in suicide (Postolache *et al.* 2005). Both of these exposures vary seasonally and there is a known biological process whereby drought stress increases allergenicity of pollen (El Kelish *et al.* 2014).

Such indicators represent forms of natural capital. They are linked to forms of human capital and these could interact with elements of the other types of capital to culminate in the risk of suicide in vulnerable sub-populations. Such an acknowledgement of the need for an holistic framework, and the development of new approaches such as the 'five-capitals', must also be cautious so that this does not suffer from the increased risk of spurious associations, multiple comparisons and practically unwieldy data collections. Adequately conveying methods and results will become even more problematic because of the multitude of steps required to conduct the analyses.

2.6 The problems of observational studies, confounders and effect modifiers

2.6.1 The 'garden of forking paths' of decisions on the path to analysis

Evidence from observational studies are given much less weight than evidence from experimental studies, however, experiments are difficult in environmental health. Observational studies face the principle problem of a large number of inter-relationships between variables. These can confound or modify effects. It is vital to a valid analysis and meaningful interpretation that we include all potential confounders and effect modifiers. Once these have been identified, scientists gather variables from a plethora of possible data sources. There is a long process of hypothesising, study design, data collection, cleaning, exploration, decision making, preparation, data analysis, model building and model checking. This process has been described as a vast 'garden of forking paths', of steps and decision points (Gelman & Loken 2013). It can become a statistical problem if scientists unconsciously or consciously introduce bias during their analysis by selecting particular data or results from a multitude of possibilities found in the course of the study. Choices made during this process may result in fallacious associations.

2.6.2 Confounding and effect modification

In studying the relationships between exposure variables and health outcomes there has been considerable work done over a long period of time to distinguish the causal roles of different variables. There are many causal factors to consider and they can interact. These variables can be assigned into one of four types: confounders, effect modifiers, causal intermediates and colliders (following defintions found in Rothman (1976) and Greenland et al. (1999)).

It is important to clearly define the scope of the study and to identify the key causal agent (or pathway of linked agents) that is the focus of the study (as discussed by Chamberlin (1890) and Platt (1964)). From that point all the other variables that may be selected for inclusion should be chosen based on their known or supposed links to the cause-effect relationship which is the object of the study. Graphical depictions of confounders, effect modifiers, causal intermediates and colliders are shown in Figure 2.4. This figure incorporates visual elements and ideas from the causal influence diagrams in Reid et al. (2012) and Greenland et al. (1999). A series of causal influence diagrams are shown as a visual example of the concepts. These concepts are crucial when attempting to use statistical modelling to measure the extent of participation of each of these causal agencies in the production of the complex result.

These graphs highlight the many causal agents involved (not only environmental but also of human biology or society). Variables that act as effect modifiers (such as the different effects by gender) are not extraneous to the effect of interest in the way that a confounder is, but are instead a component of that effect.

It is important to note that the figure does not represent the full complexity that characterizes the relations among drought, social structures, gender, allergens and suicide. Indeed its scope could be broadened in a number of ways to include other components such as temperature, humidity, air pollution, socio-economic status, unemployment, physical activity, alcohol consumption, and demographic variables such as age and gender could be introduced to the diagram. Even the mortality

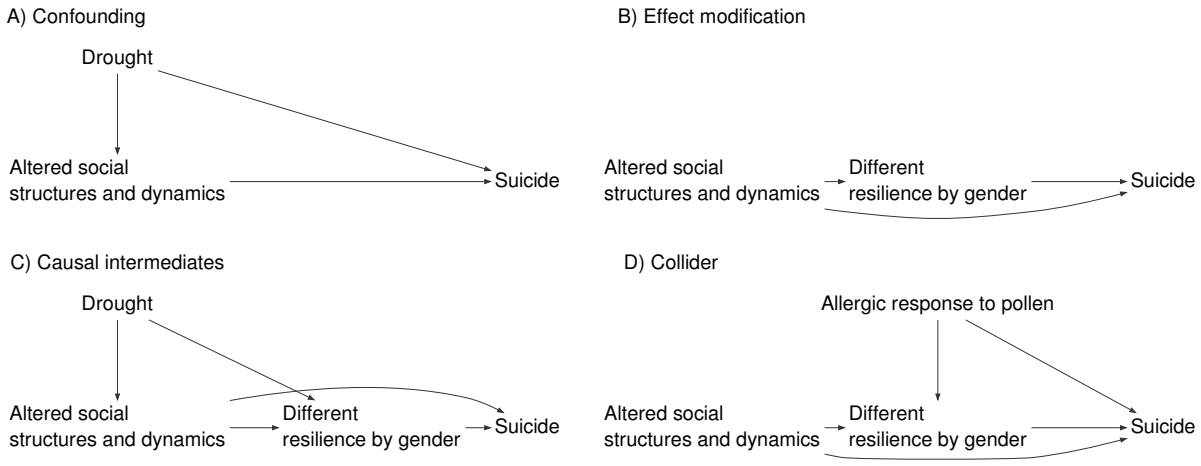


Figure 2.4: Figure depicting confounding, effect modification, intermediate and collider. Panel A displays confounding where the causal factor 'Drought' is a confounder of the relationship between 'Altered social structures and dynamics' and 'Suicide' because it is associated with both the putative cause and the outcome variable. Panel B shows an effect modifier. The 'Altered social structures and dynamics' causal factor still has direct effects on 'Suicide', but affects males and females differently because of different levels of resilience. Therefore the causal intermediary 'Different resilience by gender' has modified the effect on 'Suicide' from the social structures variable. The next example (panel C) shows that 'Drought' is an upstream cause of both 'Altered social structure and dynamics' and 'Different resilience by gender'. In this example each of the variables downstream from 'Drought' and leading to 'Suicide' would be known as a 'causal intermediate'. Panel D shows a more complicated form of confounding where an association is confounded by 'Allergic reaction to pollen', which is termed a collider. A collider is a variable that, when controlled for, biases the association between the putative cause and the outcome variable.

variable could be disaggregated into different groups based on the underlying cause of death (such as suicide by violent versus non-violent means), or co-morbidities contributing to death (such as depression).

2.6.3 Modelling effect modification

Effect modification is defined as 'differing values of the effect measure at different levels of another variate' (Rothman 1976). An example of effect modification and a multiplicative interaction model that can disentangle these causal factors (following

the description in Brambor et al. (2006)) is depicted in Figure 2.5. Further discussion of the statistical technique is provided in appendix 3: 'Decisions to be made during statistical modelling'.

Appendix 3 summarises the several options available when seeking to explain effect modification. An example is provided by paper 10 of this thesis (Vally *et al.* 2012), in which the relationship between proximity to wetlands and Ross River virus infection may be different for towns and rural areas (presumed to be due to different social and environmental factors), where the 'urban' variable is the modifier. The example shows three common ways for constructing the model parametrisation to address this question in statistical models, and the relative benefits and constraints of each are discussed in the appendix.

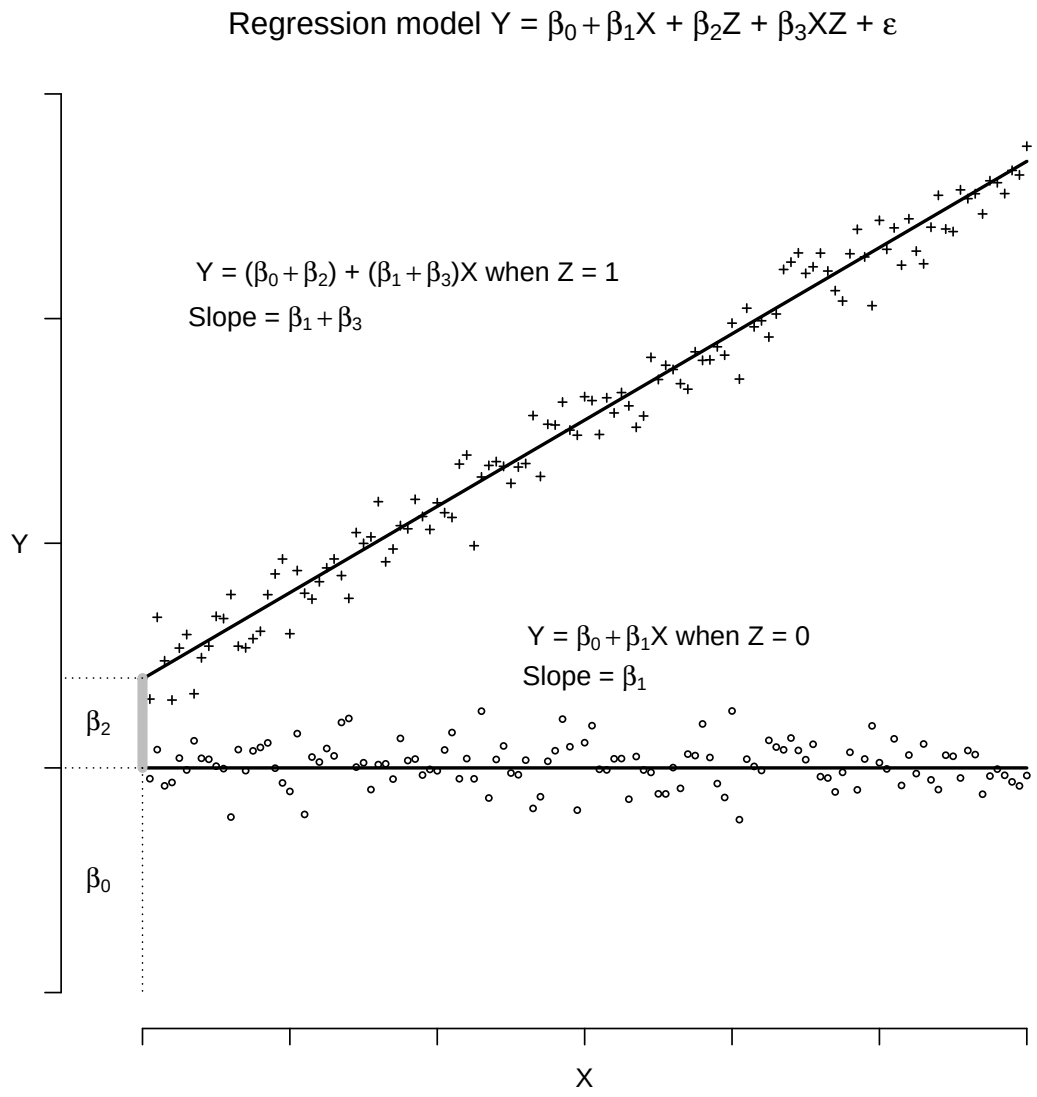


Figure 2.5: A graphical depiction of a multivariate regression model that disentangles an effect modifier

2.6.4 The primary problems of observational studies

One of the primary problems that observational studies face in demonstrating a causal link is that they must attempt to control for a great deal of inter-relationships and associations between variables. It is especially problematic that scientists select variables from a multitude of possibilities, during a process of exploration and decision making (the 'garden of forking paths' referred to above (Gelman & Loken 2013)). The terms 'researcher degrees of freedom', 'p-hacking' and 'data dredging' have also been used to add forceful cautionary messages around the misuse of statistical models of complex systems to make causal inferences that are not robust. The solution is to consider the primary concerns of making selections that are defensible to a high degree of scrutiny by a critical and sceptical reviewer. A systematic approach to the conduct of observational data analysis, following the method of 'multiple working hypotheses' (Chamberlin 1890), is conceptualised in this thesis.

These additional considerations would lead to improved understanding of the complex whole and provide more policy relevant evidence because the systemic interactions and feedbacks would be more easily identified. This will enhance our abilities to move past the simplistic identification of only proximal environmental risks. Introducing methods for using more complicated models would then lead to improvements in monitoring and evaluation of the eco-social context of human health (McMichael *et al.* 2015).

However given the nature of the problems outlined above for data analyses of many variables and disparate data types, the use of pipelines will be especially important to achieve robust scientific findings and credible policy advice. If this is not incorporated into the data analysis framework then there is a high risk of such research merely increasing the number of erroneous publications and retractions from journals.

Chapter 3

The effect on mental health of drought

3.1 Paper 1: The effect on suicide of drought in New South Wales

Title: Hanigan, I.C., Butler, C.D., Kokic, P.N. and Hutchinson, M.F. (2012). Suicide and drought in New South Wales, Australia, 1970-2007. Proceedings of the National Academy of Sciences of the United States of America, 109(35), 13950-13955. doi:10.1073/pnas.1112965109

Stage of publication: Published peer-reviewed paper

Journal Impact Factor: 9.7

My contribution: In this analysis of suicide and drought I was first author. I led the study hypothesis and design, data processing, data analysis, writing and response to reviewers. I independently created the dataset of regional suicide, population and climate data, which had never been assembled over such a long time period in Australia before. In the creation of dataset 1 and software package 1 rainfall indicators such as monthly anomalies and percentiles where created. I built on work by Prof. Hutchinson in 1992 that defined a drought index. I worked with Prof. Hutchinson to extend his method and included an additional drought index. I worked with Dr Kokic on implementation of the spatially structured time-series statistical modelling. I was responsible for the majority of the work to review relevant literature on suicide and climate/drought, I wrote the introduction and background, I independently conducted the analysis and I took the lead on submitting the manuscript and responding to reviewers.

Approvals required for inclusion of Paper 1 in this thesis:

Name	Notes	Signature/date
Mr. Ivan Hanigan		2015-09-28
Prof. Colin D. Butler		2015-10-23
Dr. Philip N. Kokic		2015-09-30
Prof. Michael F. Hutchinson		2015-10-02
Copyright	Authors retain copyright to individual articles.	2015-10-07
Ethics Approval	ANU Human Reseach Ethics Committee Protocol number: 2004/0293. Date: 01/09/2004, variation 29/09/2009 (extension and addition of personnel).	2004-09-01

Suicide and drought in New South Wales, Australia, 1970–2007

Ivan C. Hanigan[a,b,1], Colin D. Butler[a], Philip N. Kokic[c], and Michael F. Hutchinson[d]

[a]National Centre for Epidemiology and Population Health, Research School of Population Health, Australian National University, Canberra ACT 0200, Australia; [b]Commonwealth Scientific and Industrial Research Organisation, Climate Adaptation Flagship, Aspendale VIC 3195, Australia; [c]Commonwealth Scientific and Industrial Research Organisation, Mathematics, Informatics and Statistics, Canberra ACT 2601, Australia; and [d]Fenner School of Environment and Society, College of Medicine, Biology and Environment, Australian National University, Canberra ACT 0200, Australia

Edited by Kirk R. Smith, University of California, Berkeley, CA, and approved July 11, 2012 (received for review August 16, 2011)

There is concern in Australia that droughts substantially increase the incidence of suicide in rural populations, particularly among male farmers and their families. We investigated this possibility for the state of New South Wales (NSW), Australia between 1970 and 2007, analyzing data on suicides with a previously established climatic drought index. Using a generalized additive model that controlled for season, region, and long-term suicide trends, we found an increased relative risk of suicide of 15% (95% confidence interval, 8%–22%) for rural males aged 30–49 y when the drought index rose from the first quartile to the third quartile. In contrast, the risk of suicide for rural females aged >30 y declined with increased values of the drought index. We also observed an increased risk of suicide in spring and early summer. In addition there was a smaller association during unusually warm months at any time of year. The spring suicide increase is well documented in nontropical locations, although its cause is unknown. The possible increased risk of suicide during drought in rural Australia warrants public health focus and concern, as does the annual, predictable increase seen each spring and early summer. Suicide is a complex phenomenon with many interacting social, environmental, and biological causal factors. The relationship between drought and suicide is best understood using a holistic framework. Climate change projections suggest increased frequency and severity of droughts in NSW, accompanied and exacerbated by rising temperatures. Elucidating the relationships between drought and mental health will help facilitate adaptation to climate change.

self-harm | depression | rainfall | weather

Suicide, a tragic event with repercussions throughout the community, is a frequent cause of death in Australia. The Australian Bureau of Statistics reported that in 2008 suicide ranked 14th among causes of deaths registered in Australia. In recent decades, the rate has been highest in males aged 30–49 y and over 75 y (1). There is concern in Australia that the incidence of suicide is increased by drought (2). Much of rural Australia, including NSW, experiences prolonged periods of dryness and low rain. In this study, drought is defined as a persistent lack of rainfall compared with a location's median rainfall (*Materials and Methods*).

There are several plausible mechanisms by which drought may increase the suicide rate. First, droughts increase the financial stress on farmers and farming communities. Such difficulty may occur in conjunction with other economic stresses, such as rising interest rates, falling commodity prices, or an unfavorable foreign exchange rate. Second, environmental degradation can take a great psychological toll (3), which may be acute during droughts, linked with decisions and actions to sell or kill starving animals or to destroy orchards and vineyards, which in some cases were accumulated painstakingly over generations. Such loss, and even the apprehension of loss, undoubtedly places a burden on the mental health of farmers and their families. This mourning may not be confined to farmers, but also may extend to other sections of the community likely to be impoverished by long-term environmental degradation. The experience of seeing suffering wild plants and animals, or parched urban parks and gardens, and contemplation of their loss is likely to be extremely painful for some individuals. Some people may have especially high sensitivity to nature (4) and thus may be at greater risk of self-harm during drought, irrespective of whether their residence is rural or urban. However, alongside the many complex influences affecting suicide rates, we did not assume that this would be a clear signal.

To date, few studies have examined the relationship between suicide and drought. Despite the dearth of data, however, drought has been strongly linked with Australian rural suicide rates by news media and suicide prevention advocacy groups, such as the national depression initiative "Beyondblue." Attribution of suicide risk to drought is less certain than these reports suggest, and thus warrants investigation. One analysis of annual suicide rates in NSW found an association between suicide and year-to-year decline in annual rainfall between 1964 and 2001 (2). In that study, a 300-mm decrease in rainfall was associated with an ∼8% increase in suicide rate above the mean annual rate.

Another study in NSW found an association between suicide and drought over the period 1901–1998. That study focused on the association of conservative government and suicide (5). The authors argued that conservative administrations placed greater stress on individualism, diminishing social capital and enhancing the risk of suicide in vulnerable individuals. The authors controlled for drought (among other factors) and found that drought years were associated with an increase in suicide risk of ∼7% in men and 15% in women across the whole population. To our knowledge, no other studies of suicide and drought, either in or beyond Australia, have been published. However, numerous studies have explored links between suicide and climate variables other than drought. For example, temperature has been studied in many locations using various methods to define the exposure variable including daily, monthly, and annual measures (6, 7). These studies yielded conflicting results. Some found a decreased suicide rate associated with an increase in temperature (8), whereas others observed the converse (9). One study found a U-shaped response, with elevated suicide risk on extremely cold and hot days (10).

The most consistent finding possibly related to climate is a seasonal variation, with a suicide peak evident in spring. This finding is consistently observed in extratropical locations in both hemispheres (6). Several hypotheses have been proposed to explain this association, including direct and indirect climatic effects. An early theory, supporting an indirect climatic causal influence, is that social life intensifies in spring (11). A similar indirect climatic influence is implicit in the "broken promise" hypothesis, which suggests that suicide is triggered in vulnerable persons

Author contributions: I.C.H., C.D.B., P.N.K., and M.F.H. designed research; I.C.H. performed research; M.F.H. contributed new reagents/analytic tools; I.C.H., C.D.B., P.N.K., and M.F.H. analyzed data; and I.C.H., C.D.B., P.N.K., and M.F.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: ivan.hanigan@anu.edu.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1112965109/-/DCSupplemental.

when their expectations of better times following winter are unfulfilled (12). More direct climatic influences are suggested by theories that propose biochemical and neurologic changes associated with circannual rhythms, especially related to enhanced vitamin D production from greater sunlight exposure (13). Another hypothesis links increased spring pollen exposure with coincidental binge drinking, which interact to aggravate the immune response and increase suicidal depression in allergy sufferers (14).

Results

We calculated the Hutchinson Drought Index (15), which counts consecutive months of lower-than-median rainfall based on percentiles of rainfall records at each location, to investigate whether suicide risk rose with increasing duration of drought. This index has been found to successfully identify periods of declared agricultural drought. Because the Hutchinson Drought Index is skewed with many zeros, we used a logarithmic transformation. The monthly maximum temperature anomalies were not correlated with the drought index in our data (*SI Appendix*). Rainfall was not included directly in any of the models that we used.

We fitted monthly Poisson time series generalized additive models (GAMs). We expected a substantial increase in suicide associated with drought in rural men over age 30 y (assuming that this subgroup contained a high proportion of farmers), and anticipated that other age/sex/location subgroups could have different (even nonlinear) associations. We used the generalized cross-validation tool in the MGCV package of R (16) to automatically estimate the appropriate shape of subgroup response functions.

Our core model included the predictors season, age group, and sex-specific long-term suicide trends in 11 regions of NSW. We then added the drought index and temperature as the key climatic predictive variables and tested for associations with suicide. We assessed the possibility of modification of the assessed effect of the drought index on suicide owing to any of the other covariates. We used the Schwarz Bayesian information criterion (BIC) for each model to identify the best-performing model. The addition of an interaction term on sex and region improved the BIC score over the core model. The BIC scores did not show strong support for the models with the climatic predictors compared with more parsimonious models that included only season, age, sex, region, and long-term trends on time.

We then explored suicide risk in rural men by fitting interaction terms by age, sex, and geographic location. The 11 regions were classified as rural or urban based on the location of the major cities in NSW (*Materials and Methods*). We used the category of rural males age >30 y as a proxy indicator for farmers, although this likely underestimated the suicide risk in farmers, given that other researchers found a higher risk of suicide in this group compared with the surrounding rural population (17, 18).

Figs. 1 and 2 show the response functions for suicide risk and drought in rural subgroups with 95% confidence intervals (CIs). For example, the estimated relationship between drought and suicide among males aged 30–49 y is shown by a more or less straight line, rising with drought duration as measured by the drought index. A statistically significant increase in risk of 15%

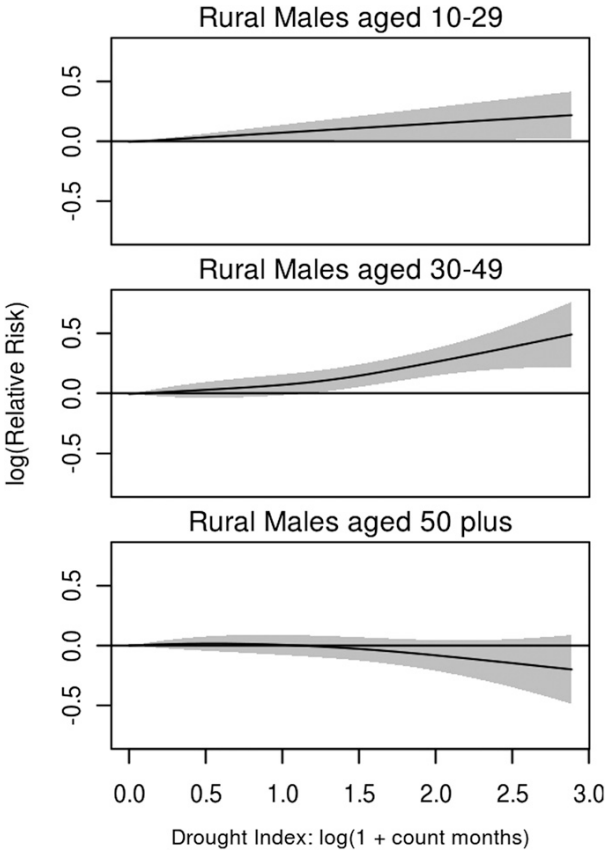


Fig. 1. Association between suicide risk and drought in rural males. Shaded areas represent 95% CIs.

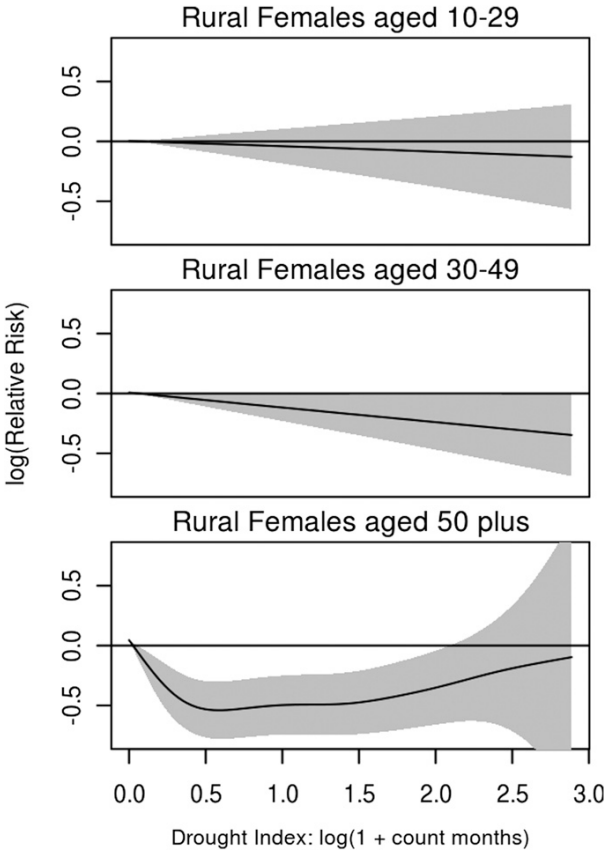


Fig. 2. Association of suicide risk and drought in rural females. Shaded areas represent 95% CIs.

(95% CI, 8–22%; $P < 0.0001$) was found when the drought index rose from the first to the third quartile.

The predicted number of annual suicides in rural males aged 30–49 associated with drought over our study period was 4.01 (95% CI, 2.14–6.05), accounting for 9% of the total number in that group over the entire 38 y of our study. Given that drought is episodic and confined to a minority of the study years, the modeled impact of drought on the number of suicides in this subpopulation is much greater than 4 per annum during drought years.

The rural female associations showed a statistically significant decrease in risk (−0.72 per annum; 95% CI, −1.32 to −0.01; $P < 0.05$). This finding was unanticipated, given that an earlier study (albeit conducted over a different time period) found a greater risk in females compared with males (5). We found that modeling the males and females separately in the rural 30–49 y age group was more statistically significant than when these were combined in the model ($P < 0.0001$, likelihood ratio test). We also found a statistically significant increase in risk associated with drought for rural males aged 10–29 y ($P < 0.01$). Our analyses for the urban population found no association between drought and suicide. Subgroup results are provided in *SI Appendix*.

Fig. 3 shows the estimated increased relative risk during spring and early summer. The relative risk of suicide increased as a linear function with warmer-than-average months. Relative risk rose by ∼3% (95% CI, 1–5%) per interquartile range rise in monthly maximum temperature anomaly (i.e., 1.6 °C).

Discussion

Our analysis of the relationship between drought and suicide was stratified by age, sex, and regional subgroups to explore different potential effects of drought, especially on farmers and farm workers. We were particularly interested in the category of rural men aged 30 y and older as a group likely to include many farmers. Our results showed an increased risk of suicide during drought in rural males aged 10–29 y and 30–49 y, but a decreased risk in rural females aged >30 y.

Although we found an increased risk due to drought in rural males aged 30–49 y, consistent with our original hypothesis that drought increases suicide risk among farmers and farming communities, the BIC rankings did not strongly support inclusion of the drought variable in the most parsimonious model (*SI Appendix*). This may be due to the occurrence of this association only in the subgroups of rural males aged 10–29 y and 30–49 y, with the relationship between drought and suicide in other subgroups (especially rural females) being different, nonexistent, or opposite. However, given our numerous model combinations, inferences based on models with less support from the BIC scores should be interpreted with caution.

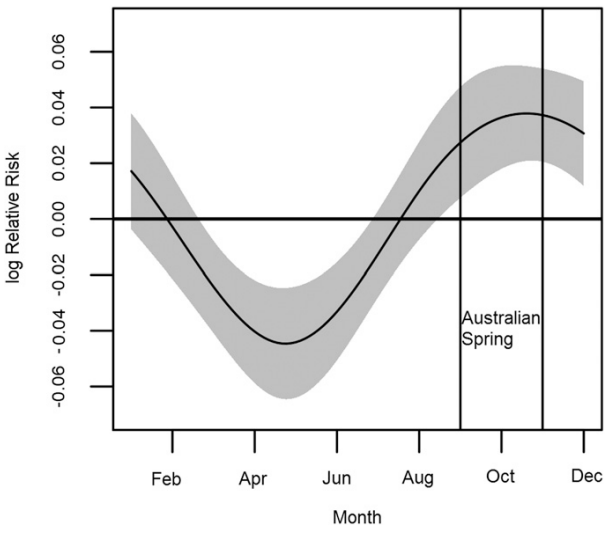


Fig. 3. Suicide risk peaks in spring and early summer. This curve is a cyclic spline derived from the core GAM (age, sex, rural location, and trend) adjusted for temperature but excluding drought.

Our results are broadly consistent with data from other studies of suicide and climate, notably that of Nicholls et al. (2), and the results in males reported by Page et al. (5). We found distinct age, sex, regional, and long-term time trends in suicide rates with increased risk associated with spring and a smaller association with above-average maximum temperatures.

The decrease in suicide risk associated with drought for rural females contradicts the data of Page et al. (5), who reported an increased risk for women during drought years. However, that study was conducted over a much longer time period with different explanatory variables. Unlike ours, it included the "sedative epidemic" of 1960–1967 (19), a period of easy availability of barbiturates, during which female suicide rates were much higher than in subsequent years.

Our modeling results do suggest that drought increases the suicide rates for males aged 30–49 in rural communities, and it is likely that many of these men are farmers or farmworkers (18). However, the risk for rural females fell. There are several possible explanations for this finding. One is that rural women may have access to more diverse social support mechanisms and thus may be able to find outlets to relieve stress. Rural women also may be more personally resilient, even in the face of drought-related hardships, including the need to care for severely depressed partners. Further, community support may strengthen, and people may pull together more as droughts persist and worsen, reinforcing social support networks in ways that particularly benefit rural women. Another possibility is that the drought declarations by the government (and associated welfare support) have a differentially beneficial influence on rural women.

Our study raises several unanswered questions. One is whether there are finer resolution regional differences, with some locations experiencing an association more than (or differently from) others. Furthermore, a marked decrease in the NSW rural suicide rates became apparent starting around 1997, even though much of NSW was in severe drought from 2000 to 2007. In 1997, in response to the Port Arthur Massacre, strict restrictions were placed on gun availability, which might have reduced suicides in general, including those associated with drought, especially in the rural community (20). The last 10 years has also seen an intensification of suicide prevention campaigns (21), increased drought support payments, and wider availability of improved antidepressant drugs. Finally, changes have been made to the way in which cause of death is coded in the database, which might have led to substantial underreporting (22).

Our study has several limitations. One is that suicide is influenced by both long-term and short-term factors, and suicidal events are likely to be lagged sometimes. The manifestation of drought's influence on suicide is likely to be complicated by these different time scales and causal pathways. We used the best available suicide and climatic data; however, the uncertain quality of the rainfall data might have reduced the precision of our calculated drought index. The drought index is based on the percentile ranking of each 6-mo average out of the entire 118 y of available rainfall record (1890–2008). Such a long period is needed to calculate extreme rainfall deficits. However, spatial rainfall models are generally considered to be of lesser quality in Australia before 1920 owing to the sparsely distributed network of monitoring stations in place at that time. In our case, however, the study region of NSW was relatively densely populated even in the 1890s. Therefore, we believe it is legitimate to include

rainfall data from the entire available period to support our analysis.

Another potential limitation is that the spatial interpolation technique used by the Bureau of Meteorology to produce the rainfall data (the Barnes inverse distance-weighted method) does not account for the influence of elevation. However, although such deficiencies would affect the magnitudes of rainfall values, they would have less effect on the percentile ranks used in the calculation of the drought index.

Suicide is a complex phenomenon with many causal influences, and thus the lack of markedly strong signals in this analysis is not surprising. Future research may benefit from the use of a more holistic framework to systematically investigate how combinations and interactions among various factors influence suicide risk. Indeed, a framework for analyzing rural livelihoods that addresses these issues does exist (23), in the form of assessment of "five capitals" comprising financial, physical, social, human, and natural forms. This conceptualization includes a more comprehensive range of explanatory variables than we have modeled in the present study. For example, we included monthly temperature anomalies as a nuisance parameter to control for potential confounding; however, it may be that the increased risk of suicide mortality is related to the fact that droughts often have higher-than-average temperatures, which can increase mortality due to thermal stress (although, as noted in *Results*, these variables are not highly correlated in our dataset).

In this framework, multiple alternate hypotheses could be explored rigorously, and the potential interplay among biological, social, and environmental mechanisms examined. Other important variables to include might be selected from different perspectives that emphasize elements from the five types of capital. For example, a biological perspective might suggest immune system function, physiological strain, and psychiatric effects of vitamin D and serotonin as indicators of human capital. An economical perspective could select farmer debt and terms of trade as indicators of financial capital. An environmental perspective might add pollen concentrations or soil degradation to drought as indicators of natural capital. A sociological perspective could identify politics and disadvantaged groups such as indigenous Australians as measures of social capital. Finally a psychological perspective could identify depressing elements of the built environment to include as indications of physical capital. Teasing apart this complex mix of causal influences is made difficult by lack of appropriate data and is beyond the scope of this paper.

It is also important to consider the various elements of environmental factors, mental health, and depression. For example, a recent study found no relationship between drought and mental health indicators (24), possibly implying an association between drought and suicide but not between drought and depression. Such an approach may help disentangle the numerous putative risk factors for rural and farmer suicide in NSW, revealing a clearer picture of the influence of drought on suicide.

Suicide and drought is an important research theme in Australia because the continent is often affected by drought. Furthermore, climate change projections indicate future increases in the frequency, intensity, and area affected by droughts in NSW, along with decreases in rainfall and humidity (25, 26). Even though these projections are less certain than those for temperature, if the rainfall changes are unexpectedly slight, then warmer temperatures (which are more confidently predicted) will exacerbate

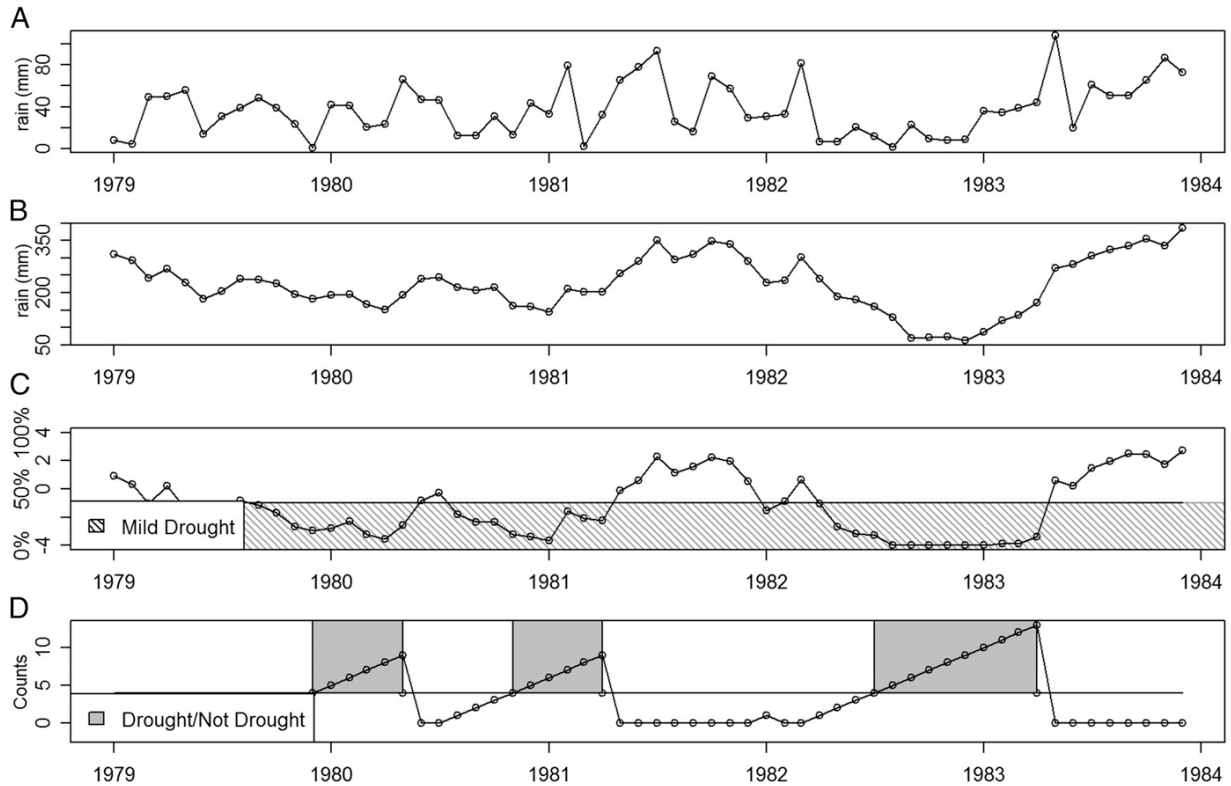


Fig. 4. Drought index in Central West NSW during a period that included a severe drought (1982–1983). The raw monthly rainfall totals (*A*) were integrated to rolling 6-mo totals (*B*), which were then ranked into percentiles by month and rescaled to range between −4 and +4 (*C*). Mild drought is below −1, and so consecutive months below this threshold were counted (*D*). A period of 5 or more consecutive months was defined as a drought in the original method.

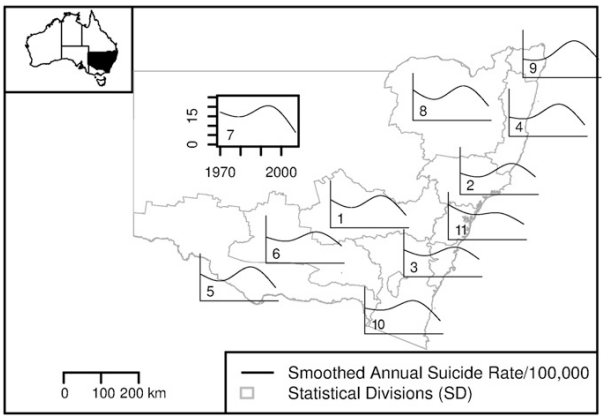


Fig. 5. NSW SD boundaries with annual suicide rate. Axes are shown on only one plot for simplicity. All axes have the same scale (*y*-axis, 0–25). Numbers refer to the SD numbers in *SI Appendix*, Tables S1 and S2.

dry periods owing to increased evapotranspiration (27). Thus, droughts are likely to increase.

Conclusions

This study addresses the substantial concern in Australia that droughts increase suicide in farmers and farm workers. We found clear evidence to support this hypothesis, with the modeled number of suicides in rural males aged 30–49 y due to drought representing ~9% of total deaths in that group over the entire 38 y of our study period 1970–2007. This subgroup is suspected to be at higher risk because of its many farmers and farmworkers. We also found an increased risk of suicide associated with drought in rural males aged 10–29 y, supporting the inference that there are flow-on effects to the broader rural community.

We investigated the data on drought and suicide at monthly and regional scales, controlling for season, age, sex, and trends over time. Using the Hutchinson Drought Index, we identified a multifaceted relationship between suicide and drought. This finding broadens the relevance of this drought index, which had previously been found to be significantly associated with declared agricultural drought.

An unexpected result was the statistically significant reduction in suicide risk during drought in rural females aged >30 y. A prominent finding was the increased risk of suicide in spring and early summer, but a lower risk of suicide in spring than during drought periods in males aged 30–49 y. However, because spring occurs every year, and because it increases the risk of suicide in all regions, for both sexes, and all age groups, the burden of suicide during spring is probably slightly greater than that of drought. We also found an association between times of unusually high maximum temperatures and increased suicide risk.

Improved understanding of these issues has important public health implications, including the timing of suicide prevention campaigns. Identifying the periods of greatest risk may allow better use of limited resources, such as promotion of counseling services to target vulnerable persons, not only during extended droughts, but also each spring. Finally, we suggest that future suicide research should consider the causation of suicide using a holistic framework that involves financial, physical, social and human factors together with natural influences, such as season and climate change.

Materials and Methods

Drought Index. We assessed the relationship between suicide and the Hutchinson Drought Index using time series Poisson GAMs of monthly data for 11 regions of NSW between 1970 and 2007. It is possible to distinguish four types of droughts: Climatic drought (based on precipitation), agricultural drought (plant and crop stress), hydrological drought (stream flow), and socioeconomic drought (supply and demand of water). Several drought indices are available. To select the index for the present study, we used three criteria: (*i*) applicability to suicide in NSW; (*ii*) ease of calculation and validity of underlying assumptions; and (*iii*) availability of spatial data to represent the extent of the droughts across the study regions. Based on these criteria, we chose the Hutchinson Drought Index (15), which integrates consecutive months of lower-than-median rainfall based on percentiles of rainfall records at each location. A full description and R codes are provided in *SI Appendix*.

The original method identified a threshold via calibration with NSW government drought declarations data to produce a binary variable for drought. However, we used the index as a continuous variable because we suspected that suicide risk would rise with increasing duration of drought. Because the drought index is skewed, we performed a logarithmic transformation. The steps taken to calculate the index are shown graphically in Fig. 4. First, the raw monthly rainfall totals (Fig. 4*A*) were integrated to rolling 6-mo total rainfall values (Fig. 4*B*) and expressed as percentiles with respect to the rainfall totals for the same 6-mo sequence over all the years of record (Fig. 4*C*). These percentiles were then linearly rescaled to lie between −4 and +4, in keeping with the range of the Palmer Index (28). Then an accounting procedure was used to integrate these percentile ranks (Fig. 4*D*). The consecutive months were counted whenever the index dropped below −1 (the threshold for mild drought on the Palmer Index). The count was reset to 0 each time the drought index rose above −1 and was restarted whenever the drought index fell below −1 again.

The best agreement with the occurrence of drought declarations for this index was found when this threshold was set at 5 mo. At this level, the optimal balance of declared droughts were successfully identified (50%) with the fewest false-positives, with a mean of 2 per drought declaration zone. As a sensitivity analysis, we assessed an enhancement to the index that increases the threshold amount of rainfall required to end a drought period from the original cutoff of −1 to a more substantial amount of 0 (i.e., the median rainfall). Descriptive statistics for the drought index are provided in *SI Appendix*, Table S1.

Ethical approval for this work was granted by the Australian National University's Human Research Ethics Committee (protocol no. 2004/0293).

Study Region. The geographic regions of the present study are areas termed Statistical Divisions (SDs) in the Australian Bureau of Statistics's population census (Fig. 5). The 11 regions were classified as rural or urban based on the locations of the three major cities of NSW: Sydney (Fig. 5, no. 11), Newcastle (Hunter SD, Fig. 5, no. 2), and Wollongong (Illawarra SD, Fig. 5, no. 3). All other SDs were classed as rural. Two SDs (North West and Far Western) were merged because their populations were considered too small to yield reliable suicide rates. All SD boundaries remained consistent from 1970 to 2007, another benefit compared with smaller areas with boundaries that change over time, confounding exposure estimation in time series studies (29).

Suicide Data. Deidentified unit records for each suicide (as determined by a coroner) in NSW between January 1970 and October 2007 were extracted from the Australian Causes of Death Unit Record File. (Data available on request from the Australian Bureau of Statistics. The Australian mortality data we use are only available subject to approval by data custodians in the government.) The final months of 2007 were excluded because of the known delay in reporting of suicide deaths. The suicide records include the day of death, age, sex, and place of usual residence of the person who died. Unfortunately, the exact location of death and time living at the place of usual residence are not recorded in the mortality database, hindering precise exposure estimation. The causes of death were coded using the International Classification of Diseases (ICD) system, which was revised three times during the study period. The codes for suicide and intentional self-inflicted injury were E950.0–E959.9 in ICD-8 (used in 1970–1978) and ICD-9 (used in 1979–1996) and X60–X84.9 and Y87.0 in ICD-10 (used from 1997 to the present). There are some issues with the comparability of ICD codes across coding system changes. For suicide deaths, ICD-8, -9, and -10 classifications are considered comparable across this period (30). We included a long-term trend variable in models that would account for changes in the time series such as this. In some other studies, injuries coded as "undetermined if accidental or intentional" are included with suicides (ICD-8 and ICD-9 codes E980.0–E989.9 and ICD-10 codes Y10.0–Y34.9, Y87.2, and Y89.9) (31–33). Over the study period, some changes to suicide coding caused some differences in the number of deaths coded as undetermined; thus, we only used these codes in a sensitivity analysis to check for the possibility of bias owing to misclassification for the other years of the study (22). The

annual suicide rates per 100,000 for each SD between 1970 and 2007 are shown in Fig. 5, and summary statistics are provided in *SI Appendix*, Table S2.

Population Data. We used the Australian census of population and housing data collected every 5 y. These data were available for local government areas for 1971–1981 and for statistical local areas for 1986–2006 (data available from the Australian Social Science Data Archives, Canberra, Australia). We assigned each local area to its SD region and categorized the populations into 10-y age groups from 10 y to 70 y and older. The population of each age and sex group in each region was linearly interpolated by month for inclusion in our models.

Climate Data. Monthly rainfall data for 1890–2008 at a resolution of 0.25 degrees latitude and longitude were used. The meteorological surfaces were constructed by the Australian Bureau of Meteorology using the Barnes inverse distance-weighted spatial interpolation approach adopted by the National Climate Centre of the Bureau of Meteorology's Research Centre in Melbourne. Monthly average maximum temperatures were obtained as well. That gridded dataset extends from 1950 to 2008. Monthly maximum temperature anomalies were calculated as the difference between each month's temperature and the long-term average for that month. The gridded monthly rainfall data were used to calculate a drought index based on 6-mo percentiles for each grid cell's entire rainfall record. These grid values were then averaged within our spatial units. We used a PostgreSQL database (http://www.postgresql.org) with the PostGIS spatial extension (http://postgis.refractions.net/) for our spatial data analysis. The Australian weather data we use are available from the Bureau of Meteorology Web site (http://www.bom.gov.au/).

Model Selection Procedure. We performed model selection using the BIC, with the model with the lowest BIC value considered the best model. Initially, we fitted a model of the death counts per month controlling for age, sex, region, season, and long-term trend in suicide rate without including climatic variables. Population was included as an offset. At first, we did not include climate variables so that we could assess the control required for our other covariates. We tested a quasi-Poisson generalized linear model to check for overdispersion. We also tested models for age-, sex-, and region-specific time trends. The trend was assumed to be unrelated to drought but to be driven by longer-term secular trends and changes, such as disease classification coding changes, antidepressant availability, or the 1997 gun control policy introduced after the Port Arthur Massacre (20). A natural cubic spline with 3 df on the sequential month rank was considered sufficient to capture these changes. Such trends may vary across age, sex, and region groups. We assessed the presence of such differences using interaction models of these variables, allowing a specific trend for each group (e.g., a sex-specific or age-specific trend), as well as three-way interactions (e.g., an age-and-sex specific trend), and finally a complete interaction as age-, sex-, and region-specific trends. There appeared to be a seasonal pattern, so we used a cyclic cubic spline with 4 df. We then added the drought term and the maximum temperature terms as penalized regression splines in GAMs, using the generalized cross-validation tool to automatically estimate the appropriate curvature of these response functions (16). Using the estimated optimal smooth on these terms, we used generalized linear models to assess all of the potential paired combination interaction models.

Analysis Code. Analyses were performed using R statistical language and environment version 2.10.0 (http://www.r-project.org). An Sweave file of the R codes used in drought index calculation and model fitting, along with additional graphs and tables, are provided in *SI Appendix*.

ACKNOWLEDGMENTS. We thank Dr. Keith Dear for statistical advice, the reviewers for their constructive suggestions that contributed to the completeness of the final version of the manuscript, and Prof. Tony McMichael and Prof. Bryan Rodgers for financial support for this work.

1. Australian Bureau of Statistics (2010) *Causes of Death, Australia, 2008* (Australian Bureau of Statistics, Canberra, Australia).
2. Nicholls N, Butler CD, Hanigan I (2006) Inter-annual rainfall variations and suicide in New South Wales, Australia, 1964-2001. *Int J Biometeorol* 50:139–143.
3. Speldewinde PC, Cook A, Davies P, Weinstein P (2009) A relationship between environmental degradation and mental health in rural Western Australia. *Health Place* 15: 865–872.
4. Louv R (2008) *Last Child in the Woods: Saving Our Children from Nature-Deficit Disorder* (Algonquin Books, Chapel Hill, NC).
5. Page A, Morrell S, Taylor R (2002) Suicide and political regime in New South Wales and Australia during the 20th century. *J Epidemiol Community Health* 56:766–772.
6. Dixon P, Kalkstein A (2009) Climate–suicide relationships: A research problem in need of geographic methods and cross-disciplinary perspectives. *Geography Compass* 3:1–14.
7. Deisenhammer EA (2003) Weather and suicide: The present state of knowledge on the association of meteorological factors with suicidal behaviour. *Acta Psychiatr Scand* 108:402–409.
8. Tsai J-F (2010) Socioeconomic factors outweigh climate in the regional difference of suicide death rate in Taiwan. *Psychiatry Res* 179:212–216.
9. Kim Y, Kim H, Kim D-S (2011) Association between daily environmental temperature and suicide mortality in Korea (2001-2005). *Psychiatry Res* 186:390–396.
10. Wang Y, Wang D, Wang X (1997) Suicide and meteorological factors in Huhhot, Inner Mongolia. *Crisis* 18:115–117.
11. Durkheim E (1951) *Suicide: A Study in Sociology* (Simon and Schuster, New York).
12. Gabennesch H (1988) When promises fail: A theory of temporal fluctuations in suicide. *Soc Forces* 67:129–145.
13. Lambert G, Reid C, Kaye D, Jennings G, Esler M (2003) Increased suicide rate in the middle-aged and its association with hours of sunlight. *Am J Psychiatry* 160:793–795.
14. Reeves GM, Tonelli LH, Anthony BJ, Postolache TT (2007) Precipitants of adolescent suicide: Possible interaction between allergic inflammation and alcohol intake. *Int J Adolesc Med Health* 19:37–43.
15. Smith DI, Hutchinson MF, McArthur RJ (1992) *Climatic and Agricultural Drought: Payments and Policy* (Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia).
16. Wood S (2008) Fast, stable direct fitting and smoothness selection for generalized additive models. *J R Stat Soc B* 70:495–518.
17. Page AN, Fragar LJ (2002) Suicide in Australian farming, 1988-1997. *Aust N Z J Psychiatry* 36:81–85.
18. Miller K, Burns C (2008) Suicides on farms in South Australia, 1997-2001. *Aust J Rural Health* 16:327–331.
19. Oliver RG, Hetzel BS (1973) An analysis of recent trends in suicide rates in Australia. *Int J Epidemiol* 2:91–101.
20. Leigh A, Neill C (2010) Do gun buybacks save lives? Evidence from panel data. *Am Law Econ Rev* 12:462–508.
21. Jorm AF, Christensen H, Griffiths KM (2006) Changes in depression awareness and attitudes in Australia: The impact of Beyondblue: the national depression initiative. *Aust N Z J Psychiatry* 40:42–46.
22. De Leo D (2007) Suicide mortality data need revision. *Med J Aust* 186:157–158.
23. Kokic P, Davidson A, Boero Rodrigez V (2006) Australia's grain industry: Factors influencing productivity growth. *Australian Commodities* 13:705–712.
24. Kelly BJ, et al. (2011) Determinants of mental health and well-being within rural and remote communities. *Soc Psychiatry Psychiatr Epidemiol* 46:1331–1342.
25. Kirono DGC, Kent DM, Hennessy KJ, Mpelasoka F (2011) Characteristics of Australian droughts under enhanced greenhouse conditions: Results from 14 global climate models. *J Arid Environ* 75:566–575.
26. Christensen JH, et al. (2007) Regional climate projections. *Climate Change 2007: The Physical Science Basis*, eds. Solomon SD, et al. (Cambridge Univ Press, New York), pp 848–940.
27. Nicholls N, Collins D (2006) Observed climate change in Australia over the past century. *Energy Environ* 17:1–12.
28. Palmer W (1965) *Meteorological Drought.* (US Department of Commerce Weather Bureau, Washington, DC), Research paper no. 45.
29. Blake M, Bell M, Rees P (2000) Creating a temporally consistent spatial framework for the analysis of inter-regional migration in Australia. *Int J Popul Geogr* 6:155–174.
30. Australian Bureau of Statistics (2000) *Suicides 1921–1998* (Australian Bureau of Statistics, Canberra, Australia).
31. Steenkamp M, Harrison J (2000) *Suicide and Hospitalised Self-Harm in Australia* (Australian Institute of Health and Welfare, Canberra, Australia).
32. Caldwell TM, Jorm AF, Dear KB (2004) Suicide and mental health in rural, remote and metropolitan areas in Australia. *Med J Aust* 181(7, Suppl):S10–S14.
33. Wilkinson D, Gunnell D (2000) Youth suicide trends in Australian metropolitan and non-metropolitan areas, 1988-1997. *Aust N Z J Psychiatry* 34:822–828.

3.1.1 Peer-reviewed supporting information

My Contribution: I was the lead author of this document. I discussed the statistical approach with Dr. Kokic and Prof. Hutchinson. I created the reproducible research report using the literate programming language Sweave, and submitted this executable file for review along with the paper. The code is hosted on Github `https://github.com/ivanhanigan/SuicideAndDroughtInNSW/blob/master/SuiDrtNSW_SupportingInfo.Rnw`

Online Supporting Information for the article: "Suicide and Drought in NSW, Australia, 1970-2007".

Ivan C. Hanigan[1,2] Colin D. Butler[1] Philip N. Kokic[2]
Michael F. Hutchinson[3]

[1]National Centre for Epidemiology and Population Health, Australian National University
[2]Commonwealth Scientific and Industrial Research Organisation
[3]Fenner School of Environment and Society, Australian National University.

Contents

1 Introduction

This document accompanies the R code at this website `https://github.com/ivanhanigan/SuicideAndDroughtInNSW` to calculate the Hutchinson Drought Index and fit the regression models for the paper 'Suicide and Drought in New South Wales (NSW), Australia, 1970-2007'. The calculation of the Drought Index is demonstrated using free data from the Australian Bureau of Meteorology. The suicide mortality data are not publicly available due to confidentiality restrictions. The R code we ran the regressions with is included but the original data are only available for authorised users approved by the Australian Bureau of Statistics and the NSW Registrar of Births Deaths and Marriages.

2 Drought Index

The R code includes a demonstration of the Hutchinson Drought Index [1]. This climatic drought index is shown graphically for a location in the 'Central West' SD of NSW in Figure 1.

Instructions for using R to download and analyse the spatial data from the Australian Bureau of Statistics (`http://www.abs.gov.au`) and the weather data from the Australian Bureau of Meteorology (`http://www.bom.gov.au`) websites are included.

2.1 Calculate the Drought Index

The Drought index is shown in Figure 1 for the SD of 'Central West NSW' during a period which includes a strong drought (1979-83). The raw monthly rainfall totals are integrated to rolling 6-monthly totals (both shown in first panel) which are then ranked into percentiles by month and this is rescaled to range between -4 and +4 in keeping with the range of the Palmer Index [2] (second panel). Mild drought is below -1 in the Palmer index and so consecutive months below this threshold are counted. In the original method 5 or more consecutive months was defined as the beginning of a drought, which continued until the rescaled percentiles exceed -1 again (third panel). The enhanced method imposes a more conservative threshold of zero (the median) to break a drought (fourth panel).

There was also an alternative method devised by Hutchinson where the rescaled percentile values are integrated using conditional cumulative sums. That method is included in the R code however we decided not to use it in this study because the counting method is simpler and gives similar results.

3 Suicide and Drought Modeling

3.1 Descriptive Statistics of Drought and Suicide

Descriptive statistics for the Drought Index are shown in Table 1. Summary statistics for Suicide rates are shown in Table 2.

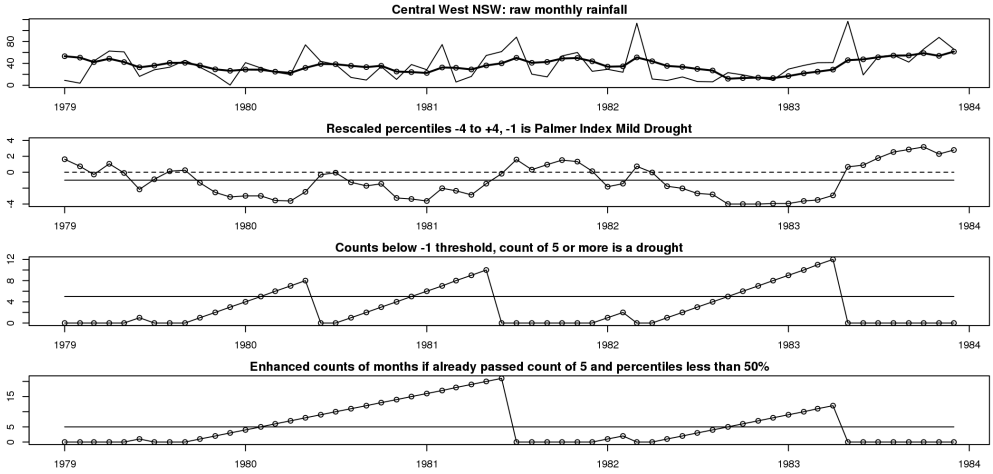


Figure 1: The Drought index in Central West NSW with the enhanced method shown in the fourth panel.

Table 1: Descriptive statistics for the drought index

SD group	N droughts	Avg Duration	Max Duration
1 Central West	9	8	12
2 Hunter	11	7	15
3 Illawarra	7	9	16
4 Mid-North Coast	8	8	15
5 Murray	7	8	11
6 Murrumbidgee	10	7	11
7 North and Far Western	8	7	12
8 Northern	5	8	11
9 Richmond-Tweed	13	8	17
10 South Eastern	8	8	11
11 Sydney	9	9	20

Table 2: Descriptive statistics for suicide (PYL = Person Years Lived)

SD group	Avg Death/Month	Avg Pop	Rate/100000 PYL
1 Central West	2	138202	13
2 Hunter	5	430403	13
3 Illawarra	3	280037	13
4 Mid-North Coast	2	183521	12
5 Murray	1	86221	14
6 Murrumbidgee	1	118778	13
7 North and Far Western	2	114460	16
8 Northern	2	146465	14
9 Richmond-Tweed	2	139356	14
10 South Eastern	2	135091	14
11 Sydney	34	3040952	13

3.2 Correlation between Temperature and Drought

We found that monthly maximum temperature variables are not strongly correlated with the drought index in our dataset. Correlation coefficients for the variables are shown in Table 3.

Table 3: Correlations

Variables	Correlation
cor(logDroughtCount,tmax)	0.05
cor(tmax,tmaxanomaly)	0.23
cor(logDroughtCount,tmaxanomaly)	0.35

3.3 Core Model Diagnostics and Variable Selection

We initially fitted age stratified time series Poisson Generalized Linear Models (GLMs). We identified a Core Model that included age, sex, region, season and long term trend. We assessed standard model diagnostics for this. Then we used Generalized Additive Models (GAMs) with the automatic estimation of the optimal amount of smoothing on the drought index using penalised regression splines from the R package: mgcv [3]. These estimated smooths were then explored in GLMs. Many models were fitted to test different combinations of variables. The models are ranked by their Bayesian Information Criterion (BIC) scores in Table 4 (AIC is shown for interest).

Table 4: Models ranked by Bayesian Information Criterion (BIC).

Model	Parameters	BIC	AIC
sd_group*sex	78	69715	69001
age*sex*ns(time,df=3)	68	69814	69191
tmaxanomModel	69	69816	69184
tmax_anomaly*sex	70	69827	69186
tmaxModel	71	69830	69180
ns(tmax,3)*sex	74	69831	69154
ageSexTrendSineXtra	70	69835	69194
droughtModel	73	69845	69186
tmax_anomaly*ns(time,3)	72	69847	69188
sd_group*ns(time,3)	98	69859	68962
With Rural 30-49 Sex Strata	98	69869	69127
Without Rural 30-49 Sex Strata	97	69873	69141
ns(tmax,3)*tmax_anomaly	75	69874	69187
tmax_anomaly*agegp	75	69877	69191
ns(logDroughtCount,5)*tmax_anomaly	79	69884	69179
ns(logDroughtCount,5)*sex	78	69885	69189
tmax_anomaly*sd_group	79	69904	69181
ns(tmax,3)*ns(time,3)	80	69909	69176
interactionDrtAgeSexRuralModel2	188	69937	69091
ns(logDroughtCount,5)*ns(time,3)	88	69963	69194
ns(logDroughtCount,5)*ns(tmax,3)	91	69968	69171
ns(tmax,3)*agegp	89	70012	69197
sd_group*sex*ns(time,3)	138	70075	68812
ns(logDroughtCount,5)*agegp	103	70080	69201
ns(tmax,3)*sd_group	101	70117	69192
ns(logDroughtCount,5)*sd_group	123	70250	69225
agegp*sd_group	128	70347	69175
sd_group*age*sex*ns(time,df=3)	618	74801	69143

3.4 Suicide and Drought Model by Age, Sex and Region

Our final GAM estimated curved response functions for drought and suicide by age, sex and region are shown in Figure 2. This model is labelled 'interactionDrtAgeSexRuralModel2' in Table 4. It included drought effects for each age/sex/region subgroup:

$$
\begin{aligned}
log(O_{ijk}) \quad = \quad & s(Drought \times Sex \times AgeGroupBy20years \times RuralOrUrbanRegion) \\
& + AgeGroupBy10years_i \times Sex_j \times s(Time, df = 3, basis = NaturalCubicSpline) \\
& + StatisticalDivision_k \\
& + s(Month, df = 4, basis = CyclicCubicSpline) \\
& + s(tmaxAnomaly) \\
& + offset(log(Pop_{ijk}))
\end{aligned}
$$

Where:

O_{ijk} = monthly suicide counts by AgeGroupBy10years$_i$, Sex$_j$ and StatisticalDivision$_k$

s(Drought × Sex × AgeGroupBy20years × RuralOrUrbanRegion) are interaction effects

Time = the month number in the sequence from Jan-1970 until Oct-2007

Month = the months of the year ranked from 1 to 12

s( ) = penalized regression splines, degrees of freedom (df) may be specified

tmaxAnomaly = monthly averaged temperature maxima anomalies from long term averages

Pop_{ijk} = interpolated population by month in each group

The eleven regions were classified as rural or urban based on the locations of the three major cities of NSW: Sydney, Newcastle and Wollongong. All other regions were classed as rural.

The estimated degrees of freedom from the GAM were then used with parametric splines in a GLM to estimate the effect sizes. A key drought effect reported in the paper was for rural males aged 30-49 where an Interquartile Range (IQR) rise in drought index gave a Relative Risk (RR) of 1.15 (95CI 1.08 to 1.22). The IQR for the drought index is about 2 months. For the temperature anomaly term there was a RR of 1.03 (95CI 1.01 to 1.05) per IQR rise (1.6 degrees C).

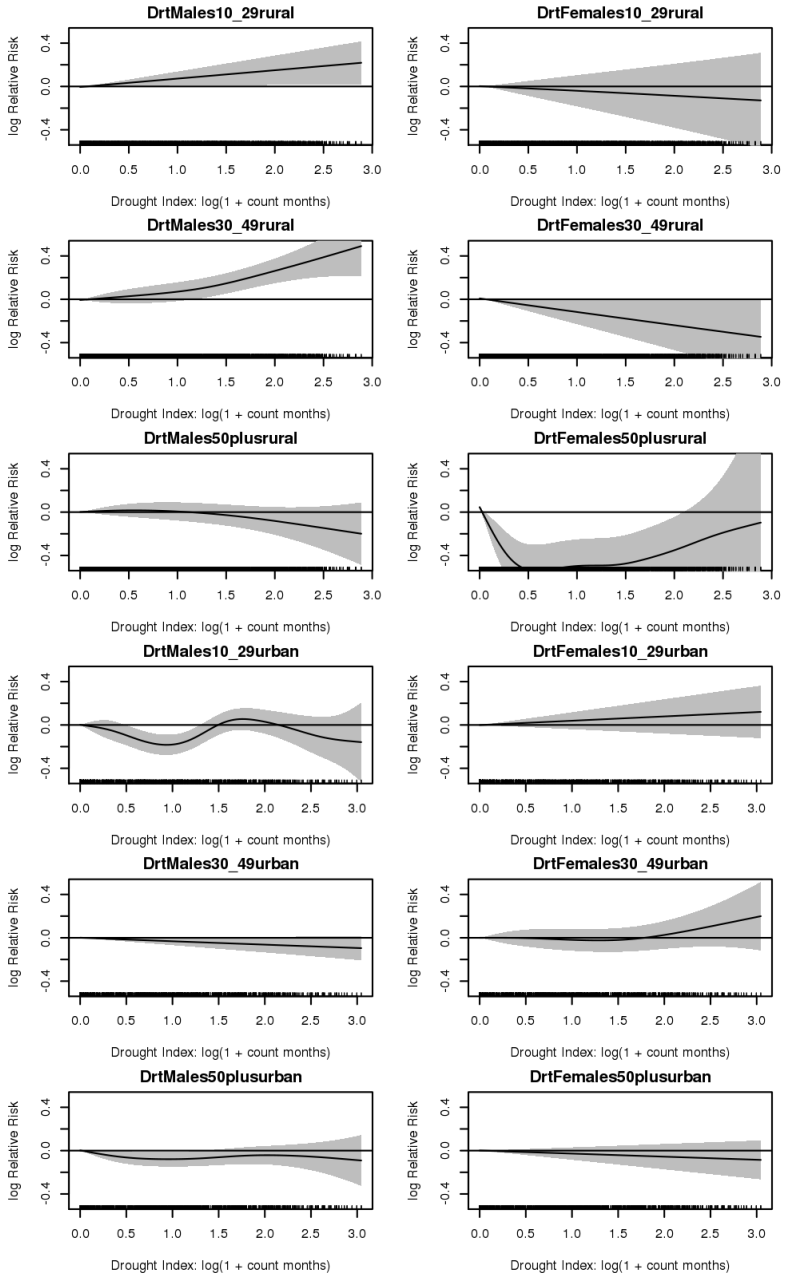


Figure 2: Estimated response functions for suicide and drought in each of the subgroups.

3.5 Attributable Number of Deaths

The predicted number of rural male suicides aged 30-49 per annum associated with droughts over our study period was 4.01 (95%CI 2.14 to 6.05, p = 0.000015), accounting for 9% of the total in 38 years.

However this effect only applies in the months that were in drought, and to a greater extent depending on the intensity of the drought. As drought is a rare and episodic event this estimate is obviously an underestimate of the real impact in terms of numbers of deaths during droughts and potential years of life lost.

The predicted number of rural female suicides aged 30-49 per annum associated with droughts are estimated for comparison with the figure for males. The decreased number of rural female suicides aged 30-49 per annum associated with droughts over our study period was -0.72 (95%CI -1.32 to -0.01, p = 0.041787).

3.6 Test the Sex Stratification

To find out if the inclusion of a seperate term for Rural Males and Rural Females aged 30-49 is warranted we performed a likelihood ratio test with an alternative model where the drought effect was not stratified by sex. The model was significantly better when including the Rural 30-49 sex stratification (likelihood ratio test p = 0.000077).

4 Sensitivity Analyses

4.1 Enhanced Drought Index

We conducted sensitivity analyses for the drought exposure variable. The drought index was enhanced with the threshold needed to end a drought made more stringent. For example in Figure 1 the drought in 1980 would not have ended in the middle of that year given the new threshold but would have continued into 1981 (the fourth panel).

The drought effects estimated were similar to those from our previous modeling.

4.2 Self-harm Coded as Undetermined

A sensitivity analysis was conducted that combined the suicide deaths with deaths coded as 'Self inflicted injury, undetermined if intentional'. This analysis agreed with our previous modelling.

4.3 Drop High Leverage Points

A sensitivity analysis was finally conducted that dropped any observations identified as having high leverage. Dropping these observations from the final model produced effect estimates that also agreed with our prior modeling results

References

[1] Smith, D. I, Hutchinson, M. F, & McArthur, R. J. (1992) *Climatic and Agricultural Drought: Payments and Policy.* (Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia).

[2] Palmer, W. (1965) *Meteorological drought. Research paper No. 45.* (U.S. Department of Commerce Weather Bureau, Washington, D.C.).

[3] Wood, S. (2008) Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 495–518.

3.1.2 Open source supporting information

My Contribution: I was the lead author of this document. This material was embedded in the reproducible research report I submitted along with the paper, however this material was not published in the journal. As it describes key additional evidence that supports the results of the analysis I published this material as open source material on my website at `https://github.com/ivanhanigan/SuicideAndDroughtInNSW/blob/master/SuiDrtNSW_SupportingInfo_OpenSource.Rnw`

Open Source Supporting Information for the article: "Suicide and Drought in NSW, Australia, 1970-2007".

Ivan C. Hanigan[1,2] Colin D. Butler[1] Philip N. Kokic[2]
Michael F. Hutchinson[3]

[1]National Centre for Epidemiology and Population Health, Australian National University
[2]Commonwealth Scientific and Industrial Research Organisation
[3]Fenner School of Environment and Society, Australian National University.

Contents

1 Introduction

This document accompanies the R code at this website `https://github.com/ivanhanigan/SuicideAndDroughtInNSW` to reproduce the work published for the paper 'Suicide and Drought in New South Wales (NSW), Australia, 1970-2007' (doi:10.1073/pnas.1112965109). These results were not included in that paper or the supporting information published with PNAS, so are published here as open source additional material.

2 Drought Index

2.1 The Summation Method

When the index is calculated using the sum of each consecutive month's rainfall deficiency score the resulting measure addresses the question of how intense the drought is, rather than just the

duration which is provided by the counting method. This version of the index is shown in Figure 1.

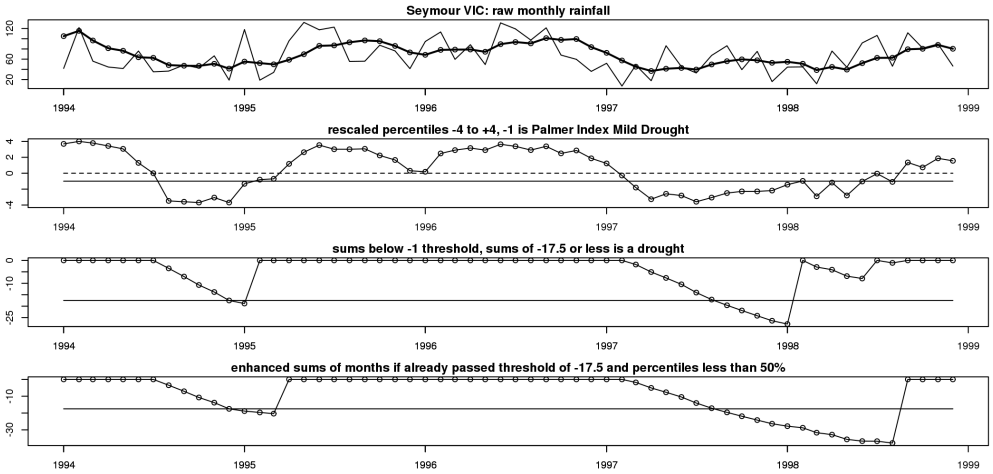


Figure 1: Drought in Seymour Victoria 1994-99 using the enhanced Hutchinson index

3 Suicide and Drought Modeling

Diagnostic plots of the core model are shown in Figure 2.

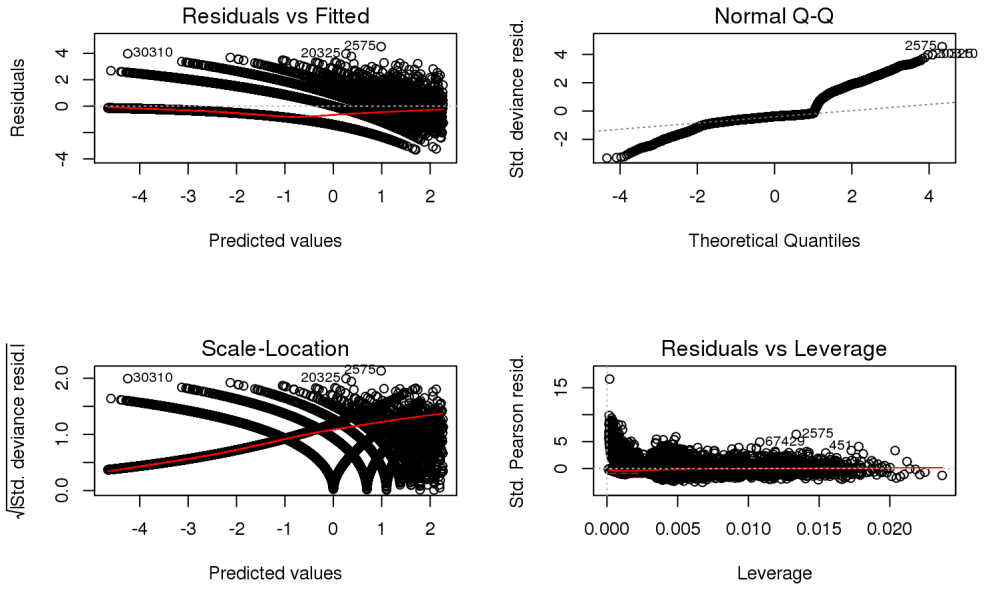


Figure 2: Core model diagnostic plots

3.1 Check Climate

The effect estimates for initial models of climate are shown in Figure 3. The drought effect was found to be complicated by the countervailing effects in men and women during subsequent modelling.

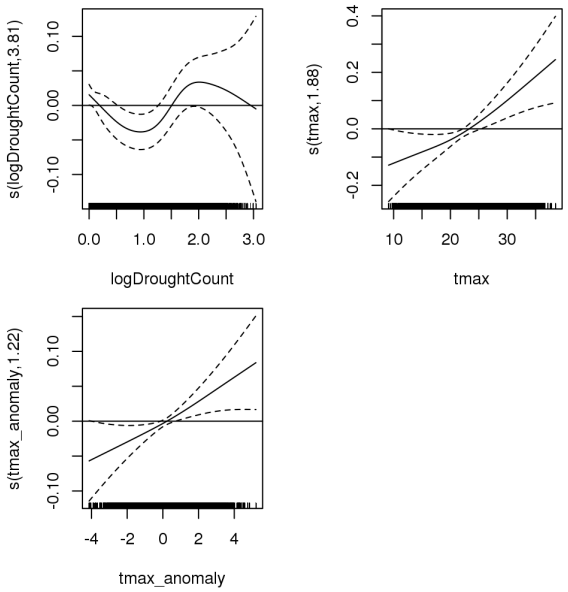


Figure 3: Drought and maximum temperature anomalies GAM results

3.2 Enhanced Drought Index

We conducted sensitivity analyses for the drought exposure variable. The drought index was enhanced with the threshold needed to end a drought made more stringent. For example in Figure 1 the drought in 1980 would not have ended in the middle of that year given the new threshold but would have continued into 1981 (the fourth panel).

The drought effects estimated were similar to those from our previous modeling.

The key effect estimates are shown for the enhanced drought index in Figure 4.

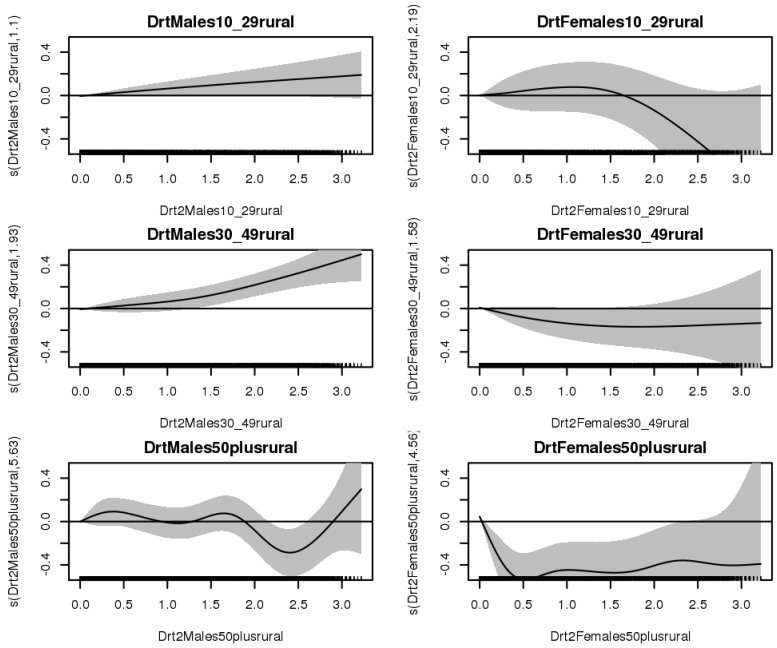


Figure 4: Interaction between drought and age, sex and rural using the enhanced Hutchinson index

3.3 Self-harm Coded as Undetermined

A sensitivity analysis was conducted that combined the suicide deaths with deaths coded as 'Self inflicted injury, undetermined if intentional'. This analysis agreed with our previous modelling.

The key effect estimates for the drought index effect on Suicides Plus Undetermined are shown in Figure 5.

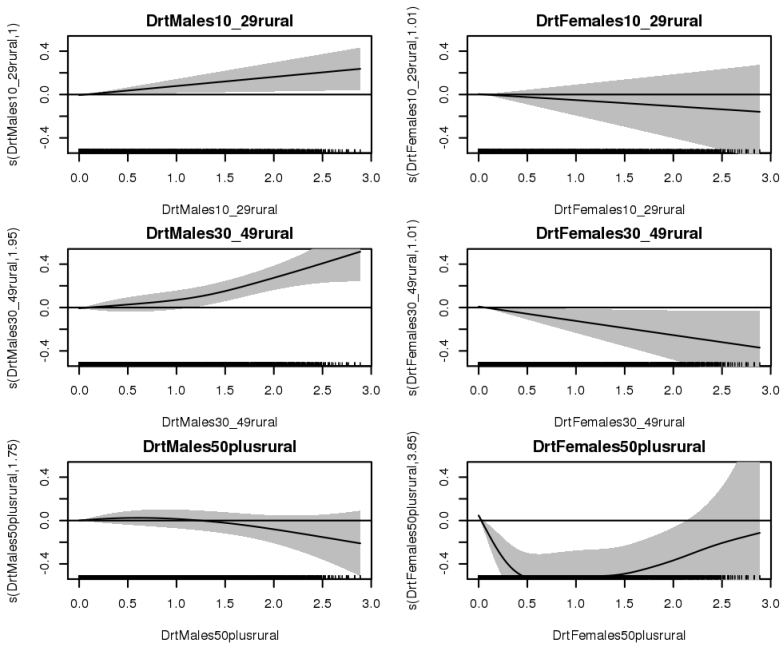


Figure 5: Interaction of drought, age, sex and rural for all suicides along with deaths with undetermined intent

3.4 Drop High Leverage Points

A sensitivity analysis was finally conducted that dropped any observations identified as having high leverage. Dropping these observations from the final model produced effect estimates that also agreed with our prior modeling results

A diagnostic plot of the leverage and residuals is shown in Figure 6. Dropping observations with high leverage produced effect estimates that also agreed with our prior modeling results (Figure 7).

3.5 Plot Check

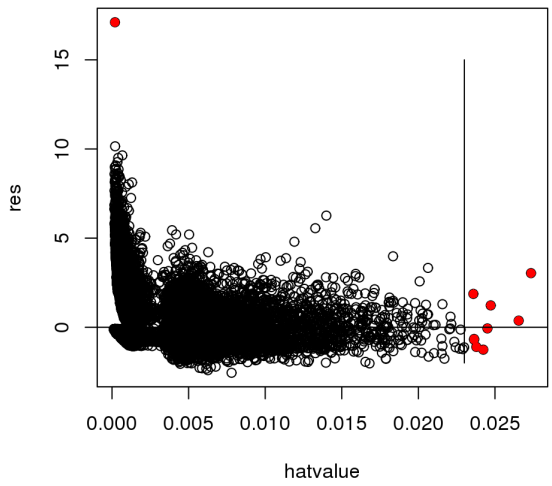


Figure 6: Interaction of drought, age, sex and rural: checks for high leverage

3.6 Plot Do

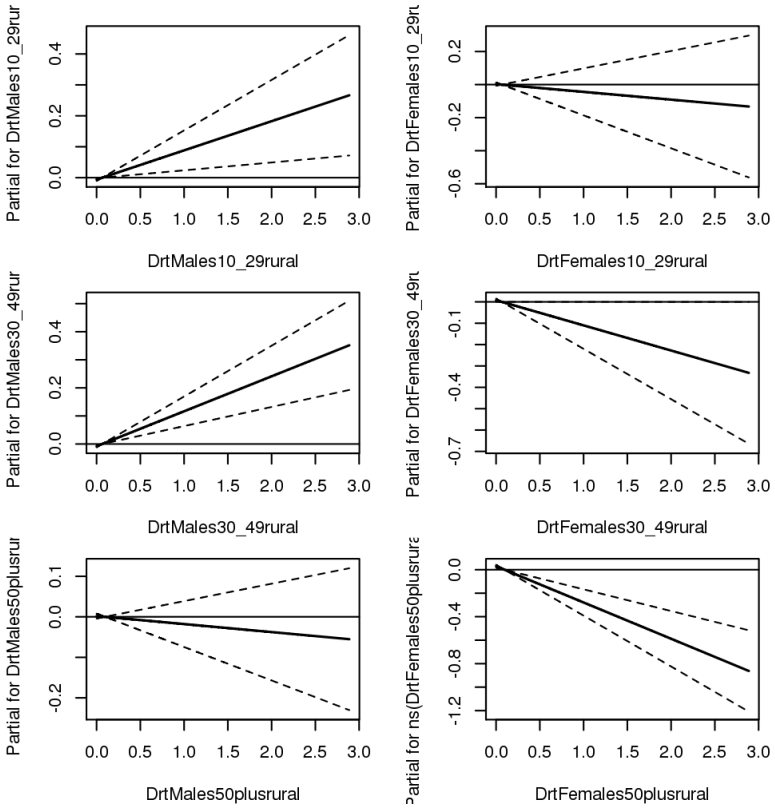


Figure 7: Interaction of drought, age, sex and rural: no leverage points

4 Code for Figures 1 and 2

The R codes to fit the model and display the exposure-response relationships for the drought index on rural suicides (Figures 1 and 2 of the paper) are shown below using Sweave.

```
> ######################
> #do,  show model fig1 and 2
> ######################
>
>
> # first fit the model
>         interactionDrtAgeSexRuralModel2 <- gam(deaths ~  s(mm, k=3, fx=T, bs = 'cp')
+         + s(DrtMales10_29rural)
+         + s(DrtMales30_49rural)
+         + s(DrtMales50plusrural)
+         + s(DrtFemales10_29rural)
+         + s(DrtFemales30_49rural)
+         + s(DrtFemales50plusrural)
+         + s(DrtMales10_29urban)
+         + s(DrtMales30_49urban)
+         + s(DrtMales50plusurban)
+         + s(DrtFemales10_29urban)
+         + s(DrtFemales30_49urban)
+         + s(DrtFemales50plusurban)
+         + s(tmax_anomaly)
+         + agegp2
+         + rural
+         + sd_group
+         + sex
+         + agegp
+         + agegp*sex*ns(time,df = 3)
+         + offset(log(pop)), data=data,family=poisson)
>
```

The code to create this graph is shown next:

```
> ######################
> #do,  show plot fig 1 and 2
> ######################
>
>
>
>         # now make a plot of each group effects
>
>         png('RuralMales20.png',res=200,width = 600, height = 1000)
>         layout(matrix(c(1:4),ncol=1),heights=c(1,1,1,0.2))
>         par(mfrow=c(4,1), mar=c(0.1,4,1.5,0.5), cex=.7)
>         plot(interactionDrtAgeSexRuralModel2,select=2,se=T, ylim = c(-0.8,0.8), shade=TRUE,shad
>         abline(0,0)
>         title('Rural Males aged 10-29', cex=.5, font.main = 1)
>         plot(interactionDrtAgeSexRuralModel2,select=3,se=T, ylim = c(-0.8,0.8), shade=TRUE,shad
>         abline(0,0)
>         title('Rural Males aged 30-49', cex=.5, font.main = 1)
>         plot(interactionDrtAgeSexRuralModel2,select=4,rug=F,se=T, ylim = c(-0.8,0.8), shade=TRU
>         abline(0,0)
>         title('Rural Males aged 50 plus', cex=.5, font.main = 1)
>         par(mar=c(1,4,6,0.5))
>         plot(1,1,type = 'n', xaxt = 'n', yaxt='n',ylab='',xlab='', axes = F)
>         title(main = 'Drought Index: log(1 + count months)',  font.main = 1,cex.main=.9)
>         dev.off()
>         png('RuralFemales20.png',res=200,width = 600, height = 1000)
>         layout(matrix(c(1:4),ncol=1),heights=c(1,1,1,0.2))
>         par(mfrow=c(4,1), mar=c(0.1,4,1.5,0.5), cex=.7)
>         plot(interactionDrtAgeSexRuralModel2,select=5,se=T, ylim = c(-0.8,0.8), shade=TRUE,shad
>         abline(0,0)
>         title('Rural Females aged 10-29', cex=.5, font.main = 1)
>         plot(interactionDrtAgeSexRuralModel2,select=6,se=T, ylim = c(-0.8,0.8), shade=TRUE,shad
>         abline(0,0)
>         title('Rural Females aged 30-49', cex=.5, font.main = 1)
>         plot(interactionDrtAgeSexRuralModel2,select=7,rug=F,se=T, ylim = c(-0.8,0.8), shade=TRU
>         abline(0,0)
>         title('Rural Females aged 50 plus', cex=.5, font.main = 1)
>         par(mar=c(1,4,6,0.5))
>         plot(1,1,type = 'n', xaxt = 'n', yaxt='n',ylab='',xlab='', axes = F)
>         title(main = 'Drought Index: log(1 + count months)',  font.main = 1,cex.main=.9)
>         dev.off()
>
>
```

3.2 Conference presentation and working paper: Open software - restricted data

Title: Hanigan, I.C., Fisher, D., and McEachern, S. (2013). Open software - restricted data: a suicide / climate case study. In National Climate Change Adaptation Research Facility Conference Climate Adaptation knowledge and partnership. June 2013, Sydney. http://swish-climate-impact-assessment.github.io/opensoftware-restricteddata/

Stage of publication: Conference poster and presentation (reviewed by the conference committee). An accompanying working paper is self published (not peer-reviewed) as open access material on my website and included in this thesis.

Journal Impact Factor: NA

My contribution: I was the lead author on this conference poster and presentation which I gave verbally. The slides and poster of the conference presentation are published on my personal website. To further demonstrate the case study I also wrote the report that is compiled with this thesis. I worked with Dr McEachern on the conceptual framework for the database and software service, including the user management and data management policy. I worked with Mr Fisher on technical aspects of the Linux server configuration and developing use cases for testing. I contributed the case study of the suicide and climate data analysis as a proof-of-concept. This database and analysis server were still in use at the National Centre for Epidemiology and Population Health in 2015. To demonstrate this case study I processed the climate change scenario data to derive projected drought index values forward one century, to the year 2100. I then demonstrated the use of the secure server when I merged the drought data with the confidential suicide data. I wrote and published the analysis code to estimate deaths attributable to drought under the two different climate scenarios.

Approvals required for inclusion of Conference presentation and working paper in this thesis:

Name	Notes	Signature/date
Mr. Ivan Hanigan		2015-09-28
Mr. David Fisher		2015-10-19
Dr. Steven McEachern		2015-10-06
Copyright	Copyright is the GNU General Public License version 2 without any warranty.	
Ethics Approval	ANU Human Reseach Ethics Committee Protocol number: 2004/0293. Date: 01/09/2004, variation 29/09/2009 extension and addition of personnel).	

Open Software - Restricted Data: A case study.

Ivan C. Hanigan [*] [1] David Fisher [2] Steven McEachern [3]

[1] *National Centre for Epidemiology and Population Health (ANU)*

[2] *Information Technology Services (ANU)*

[3] *Australian Data Archives (ANU)*

[*] *corresponding author: ivan.hanigan@anu.edu.au*

Unpublished working paper\ Draft: January 25, 2016

Background: This unpublished working paper was written to accompany the material presented as a speedtalk and poster at the National Climate Change Adaptation Research Facility Conference 'Climate Adaptation knowledge and partnership', June 2013, Sydney. The poster and slideshow are both available to download from this website: http://swish-climate-impact-assessment.github.io/opensoftware-restricteddata/presentations-nccarf-2013/. The latest version of this working paper is available to download from the website: http://swish-climate-impact-assessment.github.io/opensoftware-restricteddata/report1_high_level/manuscript.pdf.

Methods: We report on a project to build tools and procedures for enhancing open and transparent analysis of restricted datasets. Some datasets such as suicide or climate change scenarios need to be accessed in a restricted way. On the other hand scientists need to make their methods, models and assumptions transparent and available for scientific debate even though the datasets may require authorisation to access.

Results: We built a secure server/client computational environment for using open software with restricted data. We demonstrate the use of this system using drought and suicide as a case study. We describe the potential use of this system in modelling climate change scenarios.

Conclusions: The project shows that restricted data and open software can be used in an appropriate way to further the progress of scientific enquiry.

1 Background

1.1 Open software for restricted data

Some datasets such as sensitive personal information about suicide or climate change scenarios with protected intellectual property need to be accessed in a restricted way. In the context of reproducible

research methods, models and assumptions need to be made transparent and available for scientific debate even though the datasets may require authorisation to access (Peng 2011).

Restrictions around access to data have increased recently in Australia, especially to the national mortality database after the discovery of an incident in which Australian population health researcher Dr Stephen Begg was reported to have hacked into information about deaths in Australia (O'Keefe 2007). The subsequent investigations led to a wide ranging modification to the procedures for approval and provision of these data that made the access much more restricted. No new research applications were approved to access these data between 2009 and 2013.

At the same time the reproducibility crisis (Peng 2011) has emerged, reducing public confidence in statements by scientists. The true extent of the problem may turn out to be overstated (Jager & Leek 2014) however the concern should be addressed to avoid the problems that a lack of confidence in scientific publications would entail, especially in respect to evidence-based policy and expenditure of public money. Appropriate access to data and analytic software addresses this issue. We investigated available workflow tools for data management and analysis and implemented a range of these products on our server. This server has enhanced our capacity for experimentation, reviews, revisions and extensions of work in this field. We present the results of this project and report that it has streamlined access to population health and environmental data for analysis.

1.2 Motivating case study: Drought and suicide

The impact of drought on mental health is plausible, however remains a gap in epidemiological knowledge (Stanke *et al.* 2013). There is concern too that this health risk will rise under future climate change (Berry *et al.* 2008). As mentioned the number of studies that have examined the relationship between suicide and drought is limited to only a handful (Page *et al.* 2002; Nicholls *et al.* 2006; Guiney 2012; Hanigan *et al.* 2012). The motivating case for this project was to use the historical exposure-response functions to estimate future climate change impacts.

There has been substantial public interest within Australia in recent decades of the putative relationship between drought and rural mental health, including suicide. The topic has frequently been raised by the media, by rural politicians and by mental health support groups (Australian Broadcasting Commission

News 2006). There have also been media reports in India indicating substantial concerns about drought and rural suicide in that country (Sarathi Biswas 2012).

The number of studies that have examined the relationship between suicide and drought is limited. However, many papers explore links between suicide and climate variables other than drought (such as temperature) and there are two major reviews papers available of the literature on climatic influences on suicide (Deisenhammer 2003; Dixon & Kalkstein 2009). However, very few studies have investigated drought specifically.

There are several mechanisms through which unusually low rainfall, especially if exacerbated by increased soil dryness due to higher temperatures may increase the suicide rate. First, droughts increase the financial stress on farmers and farming communities (even if partially compensated by drought relief welfare payments). Such difficulty may occur in conjunction with other economic stresses, such as rising interest rates, falling commodity prices, or an unfavourable foreign exchange rate. Vins et al. (2015) provide a systematic literature review of the mental health effects of drought and explore the putative causal mechanisms.

2 Methodology

The approach we took to meet the challenge of analysing restricted suicide and climate change scenario data in a safe environment was to build a new hardware and software stack using open-source software. We based our planning on the realisation that there is a growing need of these technologies in the context of reproducible research. This requires that methods, models and assumptions need to be made transparent and available for scientific debate even though the datasets may require authorisation to access. This is true not just in health data, but also including the context of data with restrictive intellectual property and licence requirements (such as climate change scenario models).

To develop an over-all view of the issue and analyse the dimensions of the problem we spent the initial phase of the project conceptualising a rich picture of the issue, and focused on identifying risks that the project might face. Several papers that describe similar systems were reviewed (Evans & Sabel 2012; Fleming *et al.* 2014) and several recommendations from these papers were adopted in our system.

Our design responds directly to the primary threat of unintentional release of sensitive data so we decided to build a secure server/client environment for analysts to develop their software in an open way, while ensuring the safety of the datasets. Other risks we identified were in relation to the provision of the server hardware and we were able to take advantage of the Australian 'Nectar Research Cloud' (http://nectar.org.au) for virtual machines to build the servers on.

Then we defined the scope and quality of the project outcomes that we were aiming to deliver. The fact that restrictions around access to data have increased recently, coupled with arguments that appropriate access to analytic software is needed to address the reproducibility crisis meant that the scope of this project was very broad. We also explored the ambitions of our stakeholders to support their publishable outputs with open software. Given that examples of un-reproducible work has spread even to the results published in top journals (Peng 2015), the scope we decided to set for this project was for very high levels of open-ness for the evidence being presented for peer-review, along with very high levels of restriction on access to the data. Luckily however we were able to rule out the need for the extreme level of restriction such as getting Australian Defence Science and Technology Organisation (DSTO) accreditation for the security of these servers against malicious hacking attacks. Our servers just needed to be tested by the standard 'vulnerability exploits' scanner used by the Australian National University IT Department.

We also looked at the workflow system Kepler to assess it's utility for providing access to the data, but found that there were a lot of limitations at the time (Curcin & Ghanem 2008) and decided that the R environment for statistical computing and graphics would be the platform we would focus on.

During the next phase of the project we dealt with issues of the costs associated with developing the software and hosting the hardware at different locations, as well as the time needed to test and get user acceptance on the services.

2.1 System design

In this case study we utilise Virtual Machines (VMs) in the cloud. Our system requires two VMs so that the storage and processing of data can be compartmentalised, with various benefits. A high level overview of the system is shown in Figure 1. Full details including Linux commands and configuration specifications

are available online at http://swish-climate-impact-assessment.github.io/opensoftware-restricteddata/.

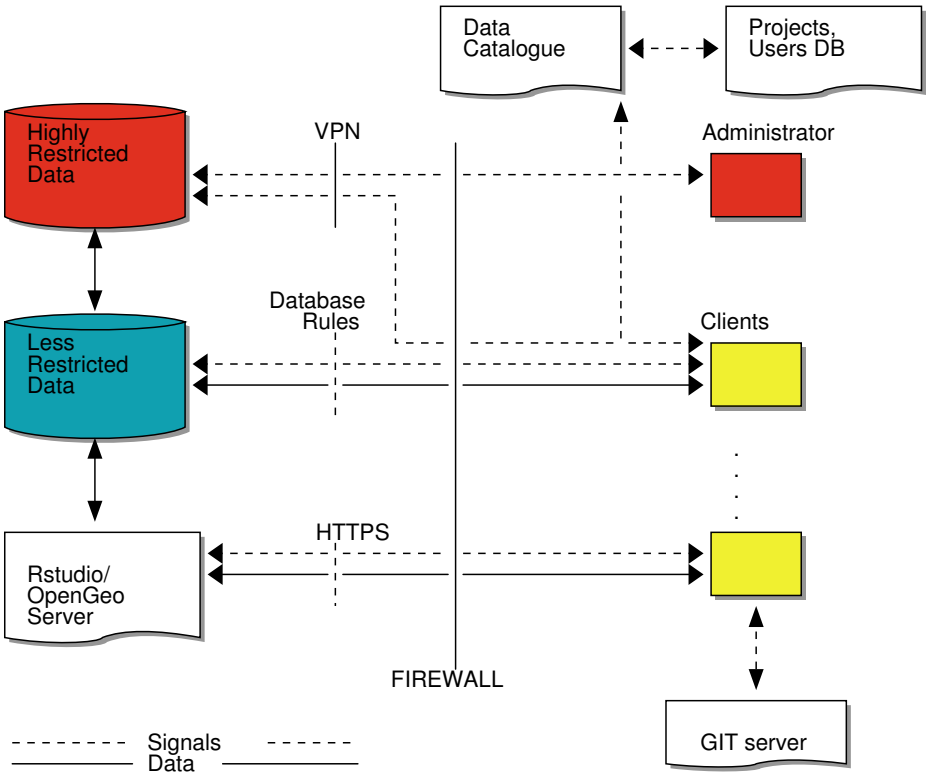


Figure 1: High Level Schematic System Design, colours indicate restrictions (red), open (blue)

2.2 Software selections

We researched a variety of systems and found the following set-up worked best for us.

Linux cluster:

- National Research Cloud www.nectar.org.au/research-cloud
- Centos 6.4 www.centos.org

Geographical Information Systems (GIS) database server:

- PostgreSQL 9.2 www.postgresql.org
- PostGIS 2.0 http://postgis.refractions.net

Statistical analysis server:

- R language for statistical computing www.r-project.org
- Rstudio server www.rstudio.com
- OpenGeo Suite http://opengeo.org

Information management:

- Projects,UsersDB Oracle XE APEX www.oracle.com
- Data Catalogue http://assda.anu.edu.au/ddiindex.html

The client side:

- Any standard web-browser
- The Kepler Project www.kepler-project.org
- pgAdmin www.pgadmin.org
- Git Version Control and GitHub www.github.com
- Emacs code editor with the starter-kit by Kieran Healy and orgmode http://kieranhealy.org/resources/emacs-starter-kit/

3 Case study

3.1 Step 1: Historical exposure-response functions

For this case study we replicated the work we had previously conducted on our personal desktop computers within the university. A key result is shown in Figure 2. That work was published already (Hanigan *et al.* 2012), however the improved IT infrastructure offered by this project allows the analysis to be re-run from a secure web-browser interface. Such improved access allowed a much broader discussion of the data, techniques and results because the researcher was able to discuss the details of the modelling with other scientists at meetings, conferences and workshops, while actually repeating the computations in real time. This is a vast improvement over the previous option of leaving the data analysis on the secure desktop computers at the university, and merely describing the computations to colleagues at the workshops.

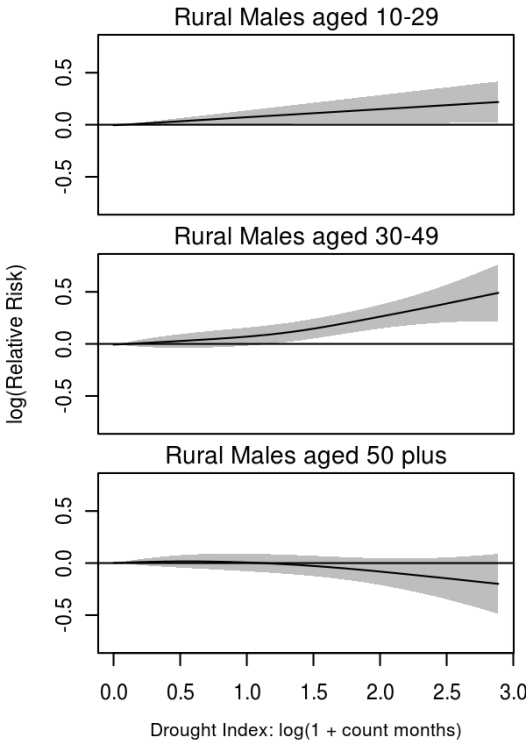


Figure 2: Drought exposure-response functions Rural Males

3.2 Step 2: Future drought scenarios and attributable fraction of suicides

Following the methods of Bambrick et al. (2008) we used the climate change scenarios provided for the Garnaut Climate Change Review (Hanigan 2013) to project estimates of future suicides under various drought conditions. The statistical method for this calculation is:

$$Y_{ijk} = \sum_{lm}(e^{(\beta_{ijk}} \times X_{lm}) - 1) \times BaselineRate_{jkl} \times Population_{jklm}$$

Where:

β_{ijk} = the ExposureVariable coefficient for zone$_i$, age$_j$ and sex$_k$

X_{lm} = Projected Future Exposure Variables

$BaselineRate_{jkl} = avgDeathsPerTime/avgPopPerTime$ in age$_j$, sex$_k$ and zone$_l$

$Population_{jklm}$ = projected populations by age_j, sex_k, $zone_l$ and $time_m$

3.3 Results

In table 1 below the two rainfall scenarios used by Berry et al. (2008) are used to demonstrate this drought impact assessment report. For full details of the data from the Garnaut Climate Change Review please refer to the original papers. For the purpose of this discussion only a brief summary is required to appreciate the relevance to this case study. The result shown below merely shows that the estimated impact of climate change can vary a lot given the input datasets and the assumptions and methods applied during analysis. The codes used to fit models, project scenarios and estimate burden of deaths are all available on the system, and can be assessed by request at the project website: http://swish-climate-impact-assessment.github.com/

Scenario	Deaths per annum	Lower 95th CI	Upper 95th percent CI
Historical (1970-2008)	4.01	2.14	6.05
A1FIR1 (Dry)	8.91	4.56	14.00
A1FIR2 (Wet)	2.93	1.50	4.47

Table 1: Estimated number of rural male suicides attributable to drought between 2000 and 2100

Below are pictorial represenations of the climate change scenarios estimated seasonal rainfall for each Statistical Division (SD) census region across the country. SDs are used because this is the geographical unit at which the suicide rates were analysed.

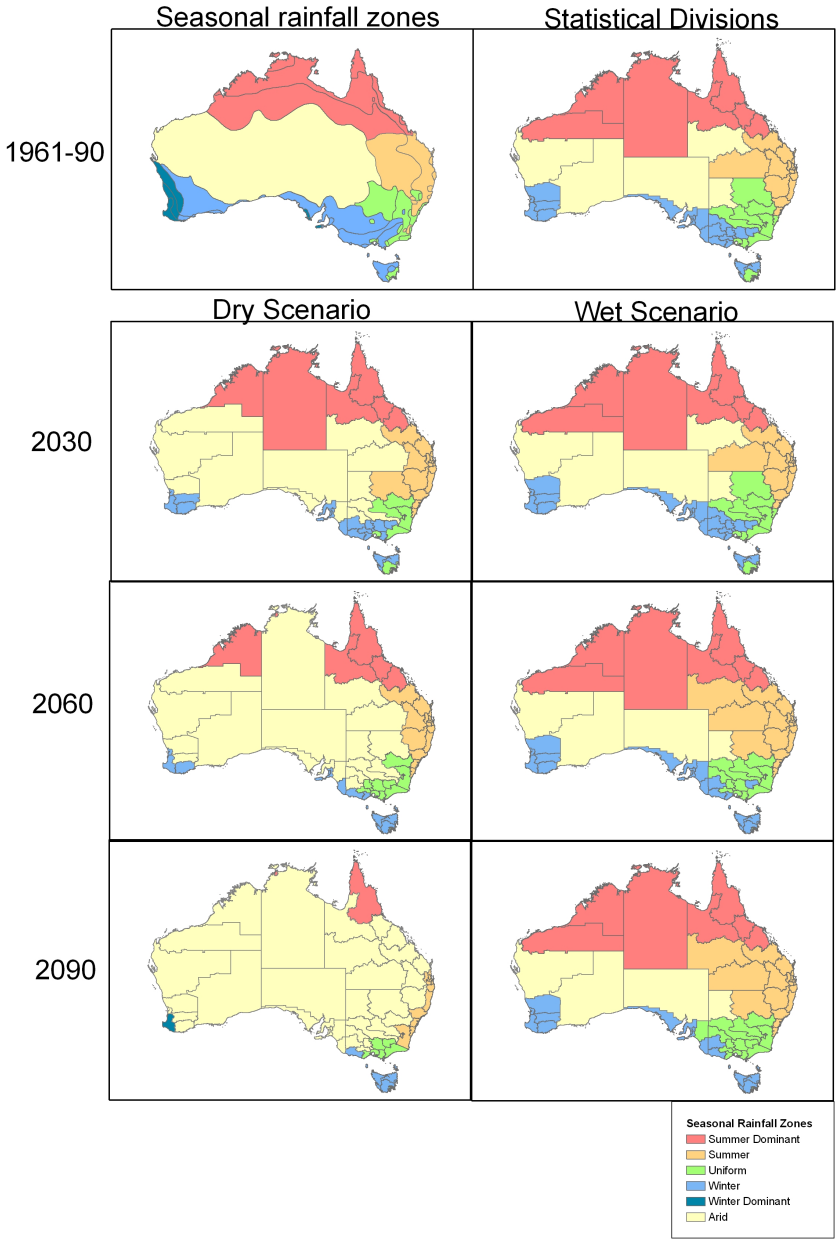


Figure 3: Seasonal rainfall under alternate future predicted scenarios (A1FI) with wet or dry assumptions

4 Discussion

4.1 Principal findings

We demonstrate the use of this system by using the climate change scenarios held on the database to project future droughts, and use the confidential suicide data that is safely stored there to estimate baseline risks and future burden of suicide attributable to the droughts. Because the server system is easy to access through a web browser these results are easily reproducible and alternate scenarios and assumptions can be tested. Because the webserver is very secure, this work can be done with confidence that the risk of unauthorised data release is minimised.

The results of our project are applicable more generally than just reproducible research. For instance in relation to issues of governance and management at the multidisciplinary institutions and multi-institutional projects it is vitally important that data sharing is enabled, while some data are still restricted because access requires approval and authorisation, or because data providers wish to be informed of re-use and identifying the profile of users of the data. It is important that data access can be made restrictive without impeding the progress of the local science agenda (collaborations, workshops, papers etc) and keeping the relevant data custodian parties informed about what is happening with the release of their datasets. The open-ness of the analytical software also has a positive effect on the value of the data infrastructure (through education and outreach activities) without risking any unethical or negligent use of these datasets.

5 Conclusions

In summary we found that the historical weather and climate change data can be attached to suicide and census data in a secure computing environment. The future drought associated deaths can then be estimated using statistical tools available in an online virtual machine. This facilitates collaboration because the data can be accessed remotely from collaborator's offices (or even workshops and conference venues) while adhering to requirements for secure data storage and access constraints.

In the case study presented it should be noted that these estimates are very uncertain, contentious and

difficult to justify. However, this reinforces the argument that data management and analysis technology such as those presented are needed to enable rigorous and transparent exploration of methods and assumptions. In this way a community of like-minded scientists can challenge or verify published results.

In conclusion the system described enables data analysis in a safe environment; allows comparison of multiple climate scenarios and assumptions; is demonstrated with a climate/health impact assessment and addresses the 'reproducibility crisis'.

6 References

Australian Broadcasting Commission News. (2006). Drought lifts suicide rates: Kennett. http://www.abc.net.au/news/2006-10-13/drought-lifts-suicide-rates-kennett/1285734 [Accessed 7 Jun. 2012]

Bambrick, H.J., Dear, K.B.G., Woodruff, R.E., Hanigan, I.C. & McMichael, A.J. (2008). *Garnaut Climate Change Review of the Impacts of Climate Change on Three Health Outcomes: temperature-Related Mortality and Hospitalisations, Salmonellosis and Other Bacterial Gastroenteritis, and Population at Risk from Dengue*. Garnaut Climate Change Review, Canberra. http://www.garnautreview.org.au/CA25734E0016A131/WebObj/03-AThreehealthoutcomes/

Berry, H.L., Kelly, B.J., Hanigan, I.C., Coates, J., McMichael, A.J., Welsh, J. & Kjellström, T. (2008). *Garnaut Climate Change Review of the Rural Mental Health Impacts of Climate Change*. http://www.garnautreport.org.au/CA25734E0016A131/WebObj/03-DMentalhealth/\protect\T1\textbraceleft/\protect\T1\textdollar\protect\T1\textbracerightFile/03-DMentalhealth.pdf

Curcin, V. & Ghanem, M. (2008). Scientific workflow systems - can one size fit all? *Cairo International Biomedical Engineering Conference*, 1–9. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4786077 [Accessed 3 Nov. 2015]

Deisenhammer, E.A. (2003). Weather and suicide: The present state of knowledge on the association of meteorological factors with suicidal behaviour. *Acta Psychiatrica Scandinavica*, 108(6), 402–409.

Dixon, P. & Kalkstein, A. (2009). Climate-suicide relationships: A research problem in need of geographic methods and cross-disciplinary perspectives. *Geography Compass*, 3(6), 1–14.

Evans, B. & Sabel, C.E. (2012). Open-Source web-based Geographical Information System for health exposure assessment. *International Journal of Health Geographics*, 11(1), 2.

Fleming, L., Haines, A., Golding, B., Kessel, A., Cichowska, A., Sabel, C., Depledge, M., Sarran, C., Osborne, N., Whitmore, C., Cocksedge, N. & Bloomfield, D. (2014). Data mashups: Potential contribution to decision support on climate change and health. *International Journal of Environmental Research and Public Health*, 11(2), 1725–1746.

Guiney, R. (2012). Farming suicides during the Victorian drought: 2001-2007. *The Australian Journal of Rural Health*, 20(1), 11–15.

Hanigan, I.C. (2013). Data for the Garnaut review of climate change 2008. https://datacommons.anu.edu.au:8443/DataCommons/rest/display/anudc:5082?layout=def:display [Accessed 7 Dec. 2015]

Hanigan, I.C., Butler, C.D., Kokic, P.N. & Hutchinson, M.F. (2012). Suicide and drought in New South Wales, Australia, 1970-2007. *Proceedings of the National Academy of Sciences*, 109(35), 13950–13955.

Jager, L.R. & Leek, J.T. (2014). An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics*, 15(1), 1–12.

Nicholls, N., Butler, C.D. & Hanigan, I. (2006). Inter-annual rainfall variations and suicide in New South Wales, Australia, 1964-2001. *International Journal of Biometeorology*, 50(3), 139–143.

O'Keefe, B. (2007). Hackers pick up UQ cash prize. *The Australian*. http://www.theaustralian.com.au/higher-education/hackers-pick-up-uq-cash-prize/story-e6frgcjx-1111113191659 [Accessed 14 Oct. 2015]

Page, A., Morrell, S. & Taylor, R. (2002). Suicide and political regime in New South Wales and Australia during the 20th century. *Journal of Epidemiology and Community Health*, 56(10), 766–772.

Peng, R.D. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226–1227.

Peng, R.D. (2015). The reproducibility crisis in science: A statistical counterattack. *Significance*, 12(3), 30–32.

Sarathi Biswas, P. (2012). Alcohol, drought lead to farmer's suicide. *Daily News and Analysis*. http://www.dnaindia.com/pune/report_alcohol-drought-lead-to-farmers-suicide_1688976 [Accessed 17 May 2012]

Stanke, C., Kerac, M., Prudhomme, C., Medlock, J. & Murray, V. (2013). Health effects of drought: A systematic review of the evidence. *PLoS Currents Disasters*, 1(1), 1–38.

Vins, H., Bell, J., Saha, S. & Hess, J. (2015). The mental health outcomes of drought: A systematic review and causal process diagram. *International Journal of Environmental Research and Public Health*, 12(10), 13251–13275.

3.3 Dataset 1: Gridded drought index data

Title: Hanigan, I. (2012). Monthly drought data for Australia 1890-2008 using the Hutchinson Drought Index. The Australian National University Australian Data Archive. doi:10.4225/13/50BBFD7E6727A

Stage of publication: Published data package at the Australian Data Archive with minted DOI

Journal Impact Factor: NA

My contribution: I was sole author of this data package. I liaised with Prof. Hutchinson to ensure the algorithm was correct and I was the corresponding author with the Australian Data Archives and the ANU Data Commons who assisted the preparation of the publication of these data. I worked on the new additions to the algorithm to implement a more conservative stopping rule to adjust the criteria for ending a drought period for each pixel.

Approvals required for inclusion of Dataset 1 in this thesis:

Name	Notes	Signature/date
Mr. Ivan Hanigan		2015-09-28
Copyright	This work is licensed under Creative Commons - Attribution 4.0. This allows others to copy, distribute and create derivative works provided that they credit the original source.	
Ethics Approval	Not applicable	

Australian National University

ANU Data Commons
Information Technology Services

- Welcome Guest
- Login

Monthly drought data for Australia 1890-2008 using the Hutchinson Drought Index

Type	Collection
Title	Monthly drought data for Australia 1890-2008 using the Hutchinson Drought Index
Brief Title	DROUGHT-BOM-GRIDS
Alternate Title	Hutchinson Drought Index
Collection Type	Collection
Access Privileges	Meteorology and Health
DOI - Digital Object Identifier	10.4225/13/50BBFD7E6727A
Metadata Language	English
Data Language	English
Significance Statement	These data were developed for the NCEPH Drought Drying and Rural Regional Mental Health collaboration.
Brief Description	The algorithm used to calculate the drought index using the Hutchinson method is stored in this repository.
Full Description	The source weather dataset are monthly spatial interpolations of rainfall data 1890-2008 from BoM weather stations, gridded at 0.25 degree of latitude-longitude across Australia. The algorithm used to calculate the drought index using the Hutchinson method is stored in this repository. The source weather data can be accessed from the principal investigators: Peter Powers and Dean Collins, National Climate Centre, BoM.
Contact Email	ivan.hanigan@anu.edu.au
Contact Address	Long Term Ecological Research Network (ANU) and Australian SuperSite Network (JCU) College of Medicine, Biology and Environment. Australian National University Canberra, ACT, 0200.
Contact Phone Number	02 6125 0349
Principal Investigator	Ivan Charles Hanigan
Supervisors	Ivan Charles Hanigan
Collaborators	Mark Clements Tony McMichael
Fields of Research	040107 - Meteorology
Keywords	Drought

Status: Published
Published To:
- Australian National Data Service
- Australian National University

Identifier: anudc:3313

Please login to request access to this dataset

Related Items

- hasCollector: http://nla.gov.au/nla.party-504967
- isOwnedBy: http://nla.gov.au/nla.party-551196
- isOwnedBy: http://nla.gov.au/nla.party-1504955
- isOwnedBy: http://nla.gov.au/nla.party-504967
- isOperatedOnBy: Extreme Weather Events Database [anudc:3314]
- hasAssociationWith: Scientific Workflow and Integration Software for Health [anudc:3315]
- hasAssociationWith: http://nla.gov.au/nla.party-504967

Files

Estimates:

- Files: 0
- Size: 0 bytes

Data Files

Date Coverage

Date From	Date To
1890	2008

Geospatial Location

Location Type	Location Value
ISO 3166-1 Country Codes	AU

Date of data creation 2012

Year of data publication 2012

Creator(s) for Citation

Given Name	Surname
Ivan	Hanigan

Publisher for Citation The Australian National University Australian Data Archive

Publications

Identifier Type
Digital Object Identifier
Identifier Value
10.1073/pnas.1112965109
Publication Title
Hanigan, IC, Butler, CD, Kokic, PN, Hutchinson, MF. Suicide and Drought in New South Wales, Australia, 1970-2007. Proceedings of the National Academy of Science USA 2012, vol. 109 no. 35 13950-13955, doi: 10.1073/pnas.1112965109.
Publication Reference

Related Websites

URL	Title
https://github.com/ivanhanigan/HutchinsonDroughtIndex/blob/master/inst/doc/HutchinsonDroughtIndex-original.pdf	Hutchinson's Drought Index
https://github.com/swish-climate-impact-assessment/DROUGHT-BOM-GRIDS	Hanigan's Github Repo
http://swish-climate-impact-assessment.github.com/EWE.html	Extreme Weather Events Database

Access Rights Contact Chief Investigator to negotiate access to the data

Licence Type CC-BY - Attribution

Data Location PostgreSQL server:
psql -h brawn.anu.edu.au -d ewedb

Example Extract:

```
select t2.*,t1.*
from bom_grids.rain_nsw_1890_2008_4 as t1 join (
select bom_grids.grid_nsw.*
from bom_grids.grid_nsw
where st_intersects(
st_GeomFromText(
'POINT('||
147.2053 ||
' '||
-33.9235 ||')'
,4283)

,bom_grids.grid_nsw.the_geom)
```

	) as t2 on t1.gid=t2.gid where year >= 2007;
Retention Period	Indefinitely
Data Management Plan	Yes

UPDATED: **13 October 2014**/RESPONSIBLE OFFICER: **Chief Information Officer** /PAGE CONTACT: **IT Services**

3.4 Software package 1: Drought index algorithm

Title: Hanigan, I., Porfirio, L. and Hutchinson, M. (2012). The Hutchinson Drought Index Algorithm. https://github.com/ivanhanigan/HutchinsonDroughtIndex.

Stage of publication: Open-source software available on Github

Journal Impact Factor: NA

My contribution: I was lead author on this software package that implements the Hutchinson Drought Index. The publication of this algorithm makes the Hutchinson Index more accessible and usable by the research community. I liaised with Dr Porfirio and Prof. Hutchinson on technical aspects of the code. I was sole author of the function to compute the Drought Indices for a point-based monitoring station. I assisted Dr Porfirio develop and test the function to compute the Drought Indices for a grid-based dataset from interpolated weather data. I led the creation of the package help manual.

Approvals required for inclusion of Software package 1 in this thesis:

Name	Notes	Signature/date
Mr. Ivan Hanigan		2015-09-28
Dr. Luciana Porfirio		2015-09-30
Prof. Michael F. Hutchinson		2015-10-02
Copyright	Copyright 2012 Ivan Hanigan and Michael Hutchinson. The program on Github is freely available under the terms of the GNU General Public License (v2). The supporting information on the Open Science Framework is licensed by CC-BY 4.0 International Public License.	
Ethics Approval	Not applicable	

Package 'HutchinsonDroughtIndex'

July 18, 2015

Type Package

Title Hutchinson's Drought Index

Version 1.0

Date 2014-01-14

Author ivanhanigan, lucianaporforio, Michael Hutchinson

Maintainer <ivan.hanigan@gmail.com>

Depends raster, rgdal, zoo

Description drought function

License GPL (>= 2)

R topics documented:

`HutchinsonDroughtIndex-package`
Hutchinson Drought Index

Description

Climatic Drought

Details

Package:	HutchinsonDroughtIndex
Type:	Package
Version:	1.0
Date:	2014-01-14
License:	GPL2

~~ An overview of how to use the package, including the most important functions ~~

Author(s)

Maintainer: Who to complain to <ivan.hanigan@gmail.com>

See Also

~~ Optional links to other man pages, e.g. ~~ ~~ <pkg> ~~

`drought_index_grids` *drought index using grids*

Description

drought index using grids

Usage

```
drought_index_grids(rasterbrick, startyear, endyear, droughtThreshold = 0.375)
```

Arguments

`rasterbrick`	a stack of grids
`startyear`	the start year
`endyear`	the end year
`droughtThreshold`	the level of dryness

`drought_index_stations` *Drought Index For Stations*

Description

Drought Index For Stations

Usage

```
drought_index_stations(data, years, droughtThreshold = 0.375)
```

Arguments

`data`	a dataframe with date, year month and rain
`years`	the number of years in the time series
`droughtThreshold`	the level of dryness below which a drought begins

Index

3.5 Dataset 2: Comparing drought estimated from a variety of data sources

Title: Hanigan, I. C. (2015). Woodland Restoration Plot Network: Climatic Drought Index Data, Western Sydney Parklands (Western Sydney Regional Park), Australia, 1887 - 2014. Long Term Ecological Research Network (LTERN). http://www.ltern.org.au/knb/metacat/ltern2.205/html

Stage of publication: Published data package at the LTERN Data Portal with persistent URL.

Journal Impact Factor: NA

My contribution: In the creation of this dataset I was the main author and conducted all data preparation and analysis. I was primarily responsible for the development of the study hypothesis, namely that the drought index values might vary depending on the type and spatial resolution of the input data. I prepared the documentation and publish the data to the portal. This dataset demonstrates the application of my prior work developing software package 1 and the dataset 1 to conduct validation assessments of the data, link these data to other research questions and to disseminate drought data through publicly available data repositories. The drought index algorithm from the R package was applied to a variety of different climatic data sources as a sensitivity analysis. The result showed that they all agreed. This was published as a dataset that gives context to drought studies at the LTERN Woodland Restoration Plot Network datasets.

Approvals required for inclusion of Dataset 2 in this thesis:

Name	Notes	Signature/date
Mr. Ivan Hanigan		2015-09-28
Copyright	These data are licensed under TERN Attribution-Share Alike-Non Commercial (TERN BY-SA-NC) Data Licence v1.0.	
Ethics Approval	Not applicable	

LTERN at TERN (http://www.tern.org.au/ltern)

LTERN Long Term Ecological Research Network (/)

HOME (/) | ABOUT LTERN (/ABOUT/WHAT-IS-LTERN) | LTERN PLOT NETWORKS (/LTERN-PLOT-NETWORKS) | DATA (/DATA)

Request an account (/data/request-account) *You are NOT logged in.* Sign in

LTERN Data Portal (http://www.ltern.org.au/knb)

Hanigan, I (2014): **Woodland Restoration Plot Network: Climatic Drought Index Data, Western Sydney Parklands (Western Sydney Regional Park), Australia, 1887 - 2014**. Long Term Ecological Research Network. http://www.ltern.org.au/knb/metacat/ltern2.205.53/html (http://www.ltern.org.au/knb/metacat/ltern2.205.53/html)

Metadata | Download EML (http://www.ltern.org.au/knb/metacat?action=read&qformat=xml&docid=ltern2.205) | Data Files 1

Identifier

DOCID **ltern2.205.53**

Data Creators

Individual	**Mr Ivan Hanigan**
Position	Data Analyst
Organization	The Australian National University
Address	Long Term Ecological Research Network, The Fenner School of Environment and Society ANU College of Medicine, Biology & Environment, The Australian National University ACTON. ACT 0200 Australia
Phone	VOICE **0428265976**
Email Address	Ivan.Hanigan@anu.edu.au (mailto:Ivan.Hanigan@anu.edu.au)

Abstract

The Woodland Restoration Plot Network Drought data package contains Hutchinson's Drought Indices calculated for the Prospect Reservoir, a location in close proximity of the Woodland Restoration Plot Network plots. The Hutchinson climatic drought indices were invented by Professor Michael Hutchinson at the Australian National University in 1992. Professor Hutchinson designed two indices to reflect the duration and the severity of agricultural droughts using precipitation data. Professor Hutchinson subsequently collaborated with Mr Ivan Hanigan at LTERN to extend the drought indices with new methods, and a computer program was written to compute the indices using a range of climatic data sources. The Woodland Restoration Plot Network Drought data package was created from three different data sources: 1) the Bureau of Meteorology Station data 1887-2014, 2) the Australian Water Availability Project's gridded climate data 1900-2014 and 3) the TERN EMAST gridded

climate data 1970-2012. The project conducted a comparative analysis of the different climatic data sources, which differ in terms of their temporal extent and spatial resolution. A synopsis of related data packages which have been collected as part of the Woodland Restoration Plot Network's full program is provided at http://www.ltern.org.au/index.php/ltern-plot-networks/woodland-restoration (http://www.ltern.org.au/index.php/ltern-plot-networks/woodland-restoration)

Contacts for Questions on the Use and Interpretation of Data

Individual	**Mr Ivan Hanigan**
Position	Data Analyst
Organization	The Australian National University
Address	Long Term Ecological Research Network, The Fenner School of Environment and Society ANU College of Medicine, Biology & Environment, The Australian National University ACTON. ACT 0200 Australia
Phone	VOICE **0428265976**
Email Address	Ivan.Hanigan@anu.edu.au (mailto:Ivan.Hanigan@anu.edu.au)

Project Information and Data Owners

Title **Woodland Restoration Plot Network**

Personnel

Role	Data Owner
Individual	**Professor David Keith**
Position	Plot Leader
Organization	Centre for Ecosystem Science, University of New South Wales
Address	Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences University of New South Wales Sydney. NSW 2052 Australia
Phone	VOICE **02 9385 2111**
Email Address	david.keith@unsw.edu.au (mailto:david.keith@unsw.edu.au)

Funding Financial support for development of the Drought Index R package was provided by Professor Tony McMichael's "Australia Fellowship" from the National Health and Medical Research Council, via the National Centre for Epidemiology and Population Health, Australian National University. LTERN funded the implementation of this method on the Woodland Restoration Plot Network site locations in 2015. This project has been part of the Long Term Ecological Research Network (LTERN). This work was supported by the Australian Government's Terrestrial Ecosystems Research Network (www.tern.org.au) – an Australian research infrastructure facility established under the National Collaborative Research Infrastructure Strategy and Education Infrastructure Fund–Super Science Initiative through the Department of Industry, Innovation, Science, Research and Tertiary Education.

Methods and Sampling Information

Methods

Method Step 1

Description

Calculation of Hutchinson's Drought Indices
The Woodland Restoration Plot Network Drought data package was created from three different data sources: 1) the Bureau of Meteorology Station data (referred to as the bom_stn source in the data package), 2) the Australian Water Availability Project's gridded climate data (labelled as awap in the data package) and 3) the TERN EMAST gridded climate data (labelled emast). The project conducted a comparative analysis of the different climatic data sources, which differ in terms of their temporal extent and spatial resolution.

The Drought indices are calculated from the monthly rainfall totals. First these are integrated to rolling 6-monthly totals which are then ranked into percentiles by month and this is rescaled to range between -4 and +4 in keeping with the range of the famous Palmer Index. Mild drought is below -1 in the Palmer index.

In the first index the duration of the drought is calculated. To do this consecutive months below the -1 threshold are counted. This is labelled duration_bom_stn when using the Bureau of Meteorology station data; duration_awap when using Australian Water Availability Project data; and duration_emast when using EMAST grid data. In the original working paper the authors calibrated their indicator to Government Drought Declarations. It was found that after 5 consecutive months a drought was often declared and so this threshold was defined as the beginning of a predicted drought declaration. This is termed the duration_xxx_declared, where xxx refers to the different data source being used to calculate the duration. This variable is set to TRUE when it is a predicted drought and FALSE otherwise. In the original method this cumulative counting step is continued until the rescaled percentiles exceed -1 again. The enhanced method implemented by Hutchinson and Hanigan imposes a more conservative threshold of zero (the median) to break a drought. It is this more stringent drought breaking criterion that is used in the Woodland Restoration Plot Network's Drought data package.

The second drought index uses a method devised by Hutchinson to calculate the severity of a drought. In this method the rescaled percentile values are integrated using conditional cumulative summation when they fall below the threshold of mild drought. This is labelled the severity_xxx variable in the data package. In the original calibration study the value of -17 was found to be the best predictor of drought declarations, and months below this value are cumulatively summed until the index value exceeds the median again (following Hutchinson and Hanigan's enhanced implementation). This is labelled the severity_xxx_duration variable.

The original 1992 working paper provides a technical description of the methods used for the original indices:

Smith, D. I, Hutchinson, M. F, & McArthur, R. J. (1992) Climatic and Agricultural Drought: Payments and Policy. (Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia).

The computer software written by Hanigan to implement the enhanced algorithm is available at: https://github.com

	/ivanhanigan/HutchinsonDroughtIndex and fully described in the paper: Hanigan, I.C, Butler, C.D, Kokic, P.N & Hutchinson, M.F. (2012). Suicide and drought in New South Wales, Australia, 1970-2007. Proceedings of the National Academy of Sciences of the United States of America, Volume: 109, Issue: 35 (2012) doi: 10.1073/pnas.1112965109 Therefore in summary the variables are defined as: • duration_bom_stn = Station Data: Consecutive months where rescaled percentiles are below the -1 threshold are counted. The count is reset when the rescaled percentiles are greater than 0. • severity_bom_stn = Station Data: Consecutive months where rescaled percentiles are below the -1 threshold are summed The summation is reset when the rescaled percentiles are greater than 0. • duration_bom_stn_declared = Station Data: After 5 consecutive months of accumulation the beginning of a predicted drought is declared. • severity_bom_stn_declared = Station Data: After the accumulated total of index values reaches -17 the beginning of a predicted drought is declared. • duration_awap = AWAP Data: Consecutive months where rescaled percentiles are below the -1 threshold are counted. The count is reset when the rescaled percentiles are greater than 0. • severity_awap = AWAP Data: Consecutive months where rescaled percentiles are below the -1 threshold are summed The summation is reset when the rescaled percentiles are greater than 0. • duration_awap_declared = AWAP Data: After 5 consecutive months of accumulation the beginning of a predicted drought is declared. • severity_awap_declared = AWAP Data: After the accumulated total of index values reaches -17 the beginning of a predicted drought is declared. • duration_emast = EMAST Data: Consecutive months where rescaled percentiles are below the -1 threshold are counted. The count is reset when the rescaled percentiles are greater than 0. • severity_emast = EMAST Data: Consecutive months where rescaled percentiles are below the -1 threshold are summed The summation is reset when the rescaled percentiles are greater than 0. • duration_emast_declared = EMAST Data: After 5 consecutive months of accumulation the beginning of a predicted drought is declared. • severity_emast_declared = EMAST Data: After the accumulated total of index values reaches -17 the beginning of a predicted drought is declared.
Instrument	None

Method Step 2	
Description	Download and process Precipitation Data for Prospect Reservoir Meteorological Station Data Monthly Precipitation Data were downloaded and processed from the Prospect Reservoir BoM station (Site name: PROSPECT RESERVOIR, Site number: 067019) http://www.bom.gov.au/climate/averages/tables/cw_067019.shtml The copyright for the source data supplied by the Bureau of Meteorology is held in the Commonwealth of Australia and this derivative work must duly give acknowledgement of the source in reference to the data.
Instrument	Site name: PROSPECT RESERVOIR Site number: 067019 Commenced: 1887 Latitude: 33.82° S Longitude: 150.91° E Elevation: 61 m

Method Step 3	
Description	Download and process Climate Grid Data: Australian Water Availability Project (AWAP) data. The middle of the reservoir was used for the gridded data analysis (2.3km away from BoM Station Location). For the pixel the AWAP grid data were downloaded and processed. The Bureau of Meteorology has generated a range of gridded meteorological datasets for Australia as a contribution to the Australian Water Availability Project (AWAP). These include monthly precipitation from 1900 to the present (http://www.bom.gov.au/jsp/awap/). Documentation is at http://www.bom.gov.au/amm/docs/2009/jones.pdf. The copyright for the source data supplied by the Bureau of Meteorology is held in the Commonwealth of Australia and this derivative work must duly give acknowledgement of the source in reference to the data.
Instrument	None

Method Step 4	
Description	Download and process Climate Grid Data: TERN Ecosystem Modelling and Scaling Infrastructure (eMAST) data. For the pixel in the middle of the reservoir the EMAST grid data were downloaded and processed. The climatic drought indices are derived from monthly total accumulated precipitation of each month, for the Australian continent between 1970-2012. The input data were modelled using ANUClimate 1.0. As this is a derivative product it must include the following citation and an email notification has been sent (9th April 2015 to eMAST.data@mq.edu.au). The citation for the source data is: Michael Hutchinson, Tingbao Xu, 2014. Monthly total precipitation: ANUClimate 1.0, 0.01 degree, Australian Coverage, 1970-2012. Australian National University, Canberra, Australia. Obtained from http://dap.nci.org.au, made available by the Ecosystem Modelling and Scaling Infrastructure (eMAST, http://www.emast.org.au) of

	the Terrestrial Ecosystem Research Network (TERN, http://www.tern.org.au). Accessed 23 March 2015 URI: http://datamgt.nci.org.au:8080/geonetwork/srv/en/metadata.show?uuid=6c8c221d-1e13-4a8f-bb91-858fa341d0d4
Instrument	None

Associated Parties

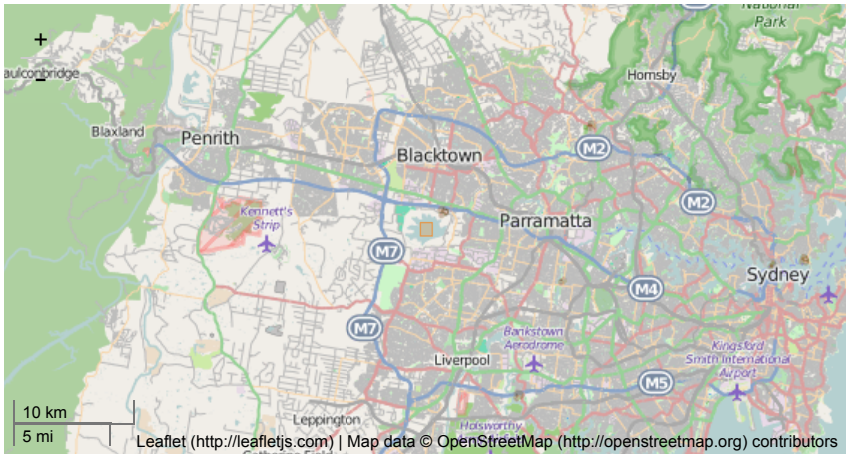
Role	Originator of the Drought Index
Individual	**Professor Michael Hutchinson**
Organization	The Australian National University
Email Address	Michael.Hutchinson@anu.edu.au (mailto:Michael.Hutchinson@anu.edu.au)

Keywords and Subject Categories

LTERN Monitoring Themes	**On plot weather** **Climate change**
Keywords List	**Rainfall data** **Precipitation** **Drought**
GCMD Science Keywords:	**Earth Science > Atmosphere > Precipitation > Rain**
ANZSRC-FOR Codes	**0599 Other Environmental Sciences**

Geographic Coverage

Geographic Description	Western Sydney Parklands (Western Sydney Regional Park), Australia
Bounding Coordinates	**WEST** **150.88792 degrees** **EAST** **150.89792 degrees** **NORTH** **-33.81607 degrees** **SOUTH** **-33.82607 degrees**



Temporal Coverage

Date Begin	1887
Date End	2014

Additional Metadata Links

3.6 Paper 2: The effect on distress of drought

Title: O'Brien, L. V., Berry, H. L., Coleman, C., and Hanigan, I. C. (2014). Drought as a mental health exposure. Environmental Research, 131: 181-187. doi:10.1016/j.envres.2014.03.014

Stage of publication: Published peer-reviewed paper

Journal Impact Factor: 4.4

My contribution: I contributed the Hutchinson Drought Index exposure data that were explored by this paper. I wrote a software tool to extract the drought index values for each pixel overlayed by the geo-coded health outcome data. This was implemented using the online restricted GIS database server which I have described in section 3.2 of this thesis (Conference presentation and working paper: 'Open software - restricted data'). I also contributed to the drafting and critical review of the manuscript. I assisted the authors to describe my original data and liaised with them as they developed their own indicators derived from them. I also gave editorial advice on descriptions of their statistical methods and results. I was responsible for providing the distribution of the associated software that was produced for this paper and distribute this software on my personal website https://github.com/ivanhanigan/DrynessAndMentalHealth.

Approvals required for inclusion of Paper 2 in this thesis:

Name	Notes	Signature/date
Dr. Lean V. O'Brien		2015-08-13
Prof. Helen L. Berry		2015-12-08
Dr. Clare Coleman		2015-09-30
Mr. Ivan Hanigan		2015-09-28
Copyright	This article is published under the terms of the CC BY NC ND licence. Permission is not required.	2015-10-07
Ethics Approval		

Environmental Research 131 (2014) 181–187

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Drought as a mental health exposure

L.V. OBrien [a,b,*], H.L. Berry [b,a], C. Coleman [b], I.C. Hanigan [a,c]

[a] National Centre for Epidemiology and Population Health, Australian National University, ACT 0200, Australia
[b] Faculty of Health, University of Canberra, ACT 2601, Australia
[c] Commonwealth Scientific and Industrial Research Organisation, CSIRO Enquiries, Clayton South, Vic. 3169, Australia

ARTICLE INFO

Article history:
Received 20 May 2013
Received in revised form
20 January 2014
Accepted 12 March 2014

Keywords:
Drought
Climate change
Mental health
Population health
Rural

ABSTRACT

The mental health impact of drought is poorly quantified and no previous research has demonstrated a relationship between distress and explicit environmentally based measures of drought. With continuing climate change, it is important to understand what drought is and how it may affect the mental health. We quantified drought in terms of duration and intensity of relative dryness and identified drought characteristics associated with poor mental health to evaluate any vulnerability in rural and urban communities. Our methods involved analysis of 100-year longitudinal records of monthly rainfall linked to one wave (2007–2008) of the Household, Income and Labour Dynamics in Australia Survey. Cluster analysis was used to characterise different patterns of dryness and linear regression analysis was used to examine associations with participant distress, as well as the moderating role of rural locality. The results showed that, during a seven-year period of major and widespread drought, one pattern of relative dryness (extreme cumulative number of months in drought culminating in a recent period of dryness lasting a year or more) was associated with increased distress for rural but not urban dwellers. The increase in distress was estimated to be 6.22%, based on 95% confidence intervals. Thus, we show that it is possible to quantitatively identify an association between patterns of drought and distress.

© 2014 Published by Elsevier Inc.

1. Introduction

Drought is common in many parts of the world and climate change projections predict it will increase substantially in frequency and/or intensity over the next century (Dai, 2010; Hennessy et al., 2008). The implications of increased drought for population mental health are potentially significant as a growing body of qualitative research indicates that drought could create a mental health exposure (Berry et al., 2010; Hossain et al., 2008; Rigby et al., 2011; Sartore et al., 2005). The clear message of this research is that mental health resilience, the capacity to cope with adversity and to avoid mental health problems when confronted by stressors (WHO, 2005), can come under unmanageable stress in the presence of prolonged drought. However, thus far, quantitative research examining the relationship between drought and mental health has been rare and piecemeal. We therefore used sophisticated methods to quantify and characterise drought in Australia and linked purpose-designed drought measures with population data taken from the Household Income and Labour Dynamics in Australia (HILDA) Survey. The method and results reported here show how drought with certain characteristics can act as a mental health exposure.

A recent review identified that the relationship between drought and mental health is an emerging area with most research being conducted in Australia (Stanke et al., 2013). The prominence of Australia is unsurprising given that Australia has the world's most variable climate and there has been a recent and prolonged drought in Australia (2001–2008). Climate change projections predict an increase in such drought events both in Australia and world-wide (Stanke et al., 2013).

Drought can be defined in several different ways but, when studying human impacts of drought, it is common practice to use the agricultural drought definition. Agricultural drought involves a period of *below-average* precipitation and/or *intense but less frequent* rain events and/or *above-normal evaporation* resulting in dry soils, reduced plant growth and reduced crop production (Dai, 2010). Implicit in this definition is the assumption that the length of an "agricultural drought" can only be assessed in hindsight because a rainfall event may signify a brief break in an extended period of dryness rather than the end of prevailing drought conditions. Consequently, in practice, researchers variously use

* Corresponding author at: Faculty of Health, University of Canberra, ACT 2601, Australia.
E-mail addresses: Lean.OBrien@canberra.edu.au (L.V. OBrien), Helen.Berry@canberra.edu.au (H.L. Berry), Clare.Coleman@canberra.edu.au (C. Coleman), Ivan.Hanigan@anu.edu.au (I.C. Hanigan).

http://dx.doi.org/10.1016/j.envres.2014.03.014
0013-9351/© 2014 Published by Elsevier Inc.

the term "drought" to mean prevailing drought conditions over time and also to refer to a specific period of dryness within the drought. For clarity, in this paper, we refer to prevailing drought conditions as "drought" and to specific periods of dryness within drought as "drought periods".

At the time of writing, the quantitative research most clearly linking drought and mental health has been focused on a specific and extreme aspect of mental health, which is suicide. Two studies using time series analysis have identified a relationship between reduced precipitation and a rise in the local suicide rate (Hanigan et al., 2012; Nicholls et al., 2006). However, of the few studies conducted using general mental health screening measures (e.g., the Kessler-10 and the SF-36), none has identified a relationship between mental health and explicit environmental measures of drought (Edwards et al., 2008; Kelly et al., 2011). This may be because these studies examined mental health in one year for the impact of low relative rainfall during a short preceding period (1–3 years), or because they did not attempt to define the nature of drought beyond identifying months of unusual dryness. In contrast, while not modelling the nature of drought, the time series suicide studies did examine trends in the impact of low relative rainfall over much longer periods known to include about a decade of prolonged drought. We therefore hypothesised that, when a drought exposure was quantified over the whole period of Australia's 2001–2008 drought (Verdon-Kidd and Kiem, 2009), it would be associated with the mental health.

In addition, although it has not previously been considered, the exact *pattern* of "drought periods" within a drought, broken by occasional rainfall events, may also influence mental health. A pattern of drought emerging over a long period of time, like the 2001–2008 drought in Australia (known as the 'Big Dry'), may take a variety of forms. Measures such as the Hutchinson indices of relative dryness can be used to provide the level of technical detail necessary to examine the emergence of these patterns. Using two different tracking methods, the Hutchinson count-method index and the Hutchinson sum-method index measure *relative* dryness across time and define benchmarks for the period of relative dryness needed before this dryness qualifies as a drought period (based on prior research examining drought in Australia: Smith et al., 1992).

Of the few studies using explicit measures of dryness, most use the Hutchinson Indices (e.g. Hanigan et al., 2012; Kelly et al., 2011; Stain et al., 2011). Although not attempted in previous studies, the Hutchinson Indices could be used to identify the number of months spent in a drought period, the duration of drought periods, how intensely "dry" the relative dryness was, and how many times an area cycled into and out of relative dryness. Measuring these characteristics would make it possible to trace how distinctive localised drought period patterns emerge within a larger region (here, a nation continent) experiencing prevailing drought conditions.

We expected that different drought period patterns (hereafter "drought patterns") would occur across the Australian continent during the 2001–2008 'Big Dry' but had no basis to make specific hypotheses about what these patterns would be or how individuals might react. We simply noted that people adapt to their environment (Hobfoll, 2002) and that mental resilience tends to break down in the face of stressors that are severe or prolonged (Pearlin, 1999). Therefore, we hypothesised that where (i) drought periods created patterns or a pattern, of *extreme* dryness conditions, that at least one of these drought patterns would create a mental health exposure, while (ii) people would adapt more easily to conditions of little or moderate drought. In addition, compared to city-dwellers, people living in rural communities live in an environmental, economic and social eco-system that is more directly connected to and reliant on the land (Connor et al., 2004). We therefore also hypothesised that rural locality would amplify any mental health impact of drought.

2. Methods

2.1. Measuring patterns of drought within the 'Big Dry' period

Extensive preparation of rainfall data was necessary to identify the drought patterns occurring in Australia during the drought of 2001–2008 known as the 'Big Dry'. The methodology used to measure dryness, drought periods and drought patterns for this study is complex and is described in detail in Appendix A of the online Supplementary materials. To briefly summarise the steps taken, we used (i) rainfall data supplied by the Australian Bureau of Meteorology to generate (ii) two 'Hutchinson Indices' of relative dryness (Hanigan, 2012). Each index used a different method of calculation (one called the 'count-method' and the other the 'sum-method') to track the dryness of different areas in Australia during the 'Big Dry'. The count method index tends to identify long drought periods while the sum method tends to identify very dry drought periods and, between them, these indices enabled us to identify when different areas entered and exited different kinds of drought conditions. (iii) Summary measures were then calculated to capture different drought characteristics experienced during the 'Big Dry'. These were: cumulative months spent in drought periods, the intensity of how relatively dry the drought period was, the number of times an area cycled into and out of a drought period, and the number of months spent in unbroken relative dryness during the last two years of the 'Big Dry'. (iv) Drought extremity measures were then calculated by dichotomising each summary measure at the 90th percentile (or as close as possible), with the exception that months spent in unbroken dryness during the last two years of the 'Big Dry' was instead dichotomised based on there being twelve straight months or more of relative dryness. Finally, (v) two-step cluster analysis in IBM SPSS 19 was used to identify the patterning of different extreme drought conditions across Australia, thereby characterising the different drought patterns Australian residents could have experienced.

Our 'drought exposure' measures were calculated to the small-area geographical level of Australian 2001 Census Collection Districts ('Districts'). The median geographical size of Districts was 32 km^2 in urban areas and 62.45 km^2 in rural areas (where the population is less dense and links to agricultural industry makes rainfall relevant across a larger geographical region). The denomination of Districts was chosen so that drought measures could be linked with population data taken from the HILDA Survey, an annual cohort survey that began in 2001 and is collected and coded in terms of Districts (details below). As is standard for large, population-level surveys, for practical reasons, HILDA survey collection is conducted over several months (for the year relevant to our study, from August 2007 to February 2008). Daily rainfall data used for calculating the drought exposure experienced by individual respondents were therefore selected to exactly match a seven year timespan that ended on the date that each participant was interviewed. Consequently, the time period examined in this study is referred to below as 2001/02–2007/08.

2.2. The Household, Income and Labour Dynamics in Australia (HILDA) Survey

2.2.1. Survey details

The HILDA Survey is a government-funded panel survey of Australians aged 15 years and over that began in 2001 (Wave 1, $N=13{,}969$). Participants provide informed consent and, for those participants who are under the age of 18 (6.98% of the current sample), a legal guardian also provides consent for their participation. The participation rates from year to year within the HILDA survey are consistently high (86% retention in the second year and above 90% in all years thereafter). For details of the sampling strategy and conduct of data collection see Appendix B in online Supplementary materials.

The HILDA Survey is conducted annually (Wooden et al., 2002). It is designed to be nationally representative with the exception that people in remote and very remote areas of Australia are underrepresented because they were not included in the initial data collection (though some participants have since moved to remote areas). The survey asks respondents to report every year on a range of aspects of life, including labour market and family dynamics, wellbeing and socio-economic characteristics. This study used Wave 7 of the HILDA Survey, Wave 7 had a response rate of 94.7% and included 13,590 respondents of whom we selected 5012 who had been resident in the same location for the past seven years (so that their weather exposure was stable) *and* had completed the mental health outcome measure. To classify participants as either urban or rural dwellers we used the 'Section of State' Australian Standard Geographical Classification system devised by the Australian Bureau of Statistics (ABS, 2011). This system uses population counts to define urban centres as those with 1000 or more people in a community while the remainder of the population is classified as rural. Within our sample, 4093 participants dwelt in an urban area and 919 dwelt in a rural area.

Wave 7 is not the most recent wave available and was selected because its collection occurred just before the end of the 'Big Dry', allowing us to capture

a population *before* it had exited or recovered from prolonged drought conditions. Further, unlike other waves, Wave 7 included Kessler's 10-item measure of general psychological distress (Kessler et al., 2002).

2.2.2. Outcome measure

The Kessler-10 is a short measure of non-specific psychological distress based on questions about nervousness, agitation, psychological fatigue and depression. It was designed be used in the United States National Health Interview Survey, and to span the range from minimal to extreme levels of distress. It has since been shown to be a valid and reliable screening tool for the Australian population with population norms available (Slade et al., 2011). Participants respond to each item in the Kessler-10 (e.g. in the past four weeks how often have you felt so sad that nothing could cheer you up?) on a five-point scale (1=none of the time, 5=all of the time), with a total summed score ranging from 10 to 50. To classify the level of distress that a score indicates, the Australian Government's Bureau of Statistics recommends the following categories: low distress=10–15; moderate distress=16–21; high distress=22–30; very high distress=30+. Those with scores of 20 or above are likely to have a mental disorder requiring clinical treatment and those with scores in the subclinical 16–19 range have an increased risk of developing a mental disorder (Coombs, 2005; Cuijpers and Smit, 2004). Scores in the present sample ranged from 10 to 50, with mean scores (M=15.53, SE=0.13), slightly higher than in the Australian population the same year (M=14.5, SE=1, Slade et al. (2011)).

2.2.3. Confounders

As standard confounders for population health-related studies, we controlled for age, sex, relationship status and gross household income adjusted for the number of people in the household using the OECD method (household income/($\sqrt{}$no. in household)). Several other confounders were also included based on research showing that they are associated with mental health (Berry et al., 2008): ethnicity; left school early (at 15 years or younger), low educational attainment, divorced/separated, receiving $>$30% of income from welfare, presence of a physical or mental health condition (either or both for 6 months or more) and satisfaction with local community.

2.2.4. Missing data

Once the HILDA Survey data were restricted to participants who had completed the dependent variable measure (Kessler 10), there were only five participant cases with missing data. The missing data was on four covariates: education, health condition, ethnicity and satisfaction with local community and there was less than 1% of missing data on each variable. To avoid loss of power, missing data were imputed using an iterative regression-based approach called Expectation Maximisation (Enders, 2001). Expectation Maximisation uses a two-step iterative process where missing values are replaced with their best estimates based on the complete data. The first step (the 'E Step'), replaces missing values with the conditional expected value given the observed data. The second step (the 'M Step') uses the estimations from the first step to produce means and covariance matrices as if there were no missing data. The mean scores and standard deviations for all the variables before and after estimation are then compared and, if the scores are sufficiently similar, the estimations can be accepted. The algorithm is repeated until the solution reaches convergence. Note that sensitivity tests showed that reducing the sample to cases with no imputed data did not change the pattern of results or introduce overlap between confidence intervals. All analyses were therefore conducted using the dataset with imputation.

2.3. Statistical analysis

Drought patterns were identified by performing cluster analyses on the drought extremity measures. Next, the drought patterns were used to predict mental health, via multiple linear regression analysis, with the five drought pattern categories coded as five categorical (dummy) variables and zero-to-moderate drought as the reference category. To examine whether rural locality would amplify any mental health impact of drought, interaction terms between drought pattern category and rural/urban locality were included. All analyses adjusted for the socio-demographic confounders and the confounders known to be related to mental health as listed in the confounders subsection above. Analyses were conducted using Stata 12 survey commands, applying the HILDA Survey sample weights and clustering at the District level, which also accounts for clustering within households. For more detail on the complex survey techniques see Appendix B in Supplementary materials.

3. Results

3.1. Patterns of dryness

Repeated cluster analyses using the measures of extreme cumulative dryness (Table 1) showed that, while drought period *cycles* were not consistently informative, a highly stable cluster solution was returned when three indicators of extreme conditions were used: (i) an extreme number of months qualifying as drought periods, (ii) extremely dry drought periods and (iii) the presence of a recent long unbroken dry period. During the 'Big Dry' these characteristics were distributed across the population in a way that created five distinct drought patterns, termed: (1) zero-to-moderate drought, (2) very dry drought, (3) recent long period, (4) constant drought, and (5) constant drought with a recent long period. See Table 1 for further description of these drought patterns, and the bottom of Table 2 for drought pattern distribution and the means of related drought extremity measures split by urban/rural locality.

Note that the total number of months spent in drought was similar across the 'constant' drought and the 'constant drought with a recent long period' categories. We also reiterate that, except for the zero-to-moderate drought classification, the drought patterns identified were focused on *extreme* dryness, or dryness over and above an allowance for normal variability (which, in Australia, is substantial). Thus, for example, a 'recent long dry period' typically signalled around 16 months in unbroken dryness during the past 24 months.

3.2. Socio-demographic characteristics

Descriptive characteristics for rural and urban dwellers are reported in Table 2. The rural population tended to be slightly more clustered around early retirement age, and had a greater proportion of men. Rural dwellers were more likely to have left school early and were less likely to have completed a university degree. They tended to have lower adjusted household income and were also slightly more likely to obtain more than 30% of their income from government pensions and allowances. There were also markedly fewer non-English-speaking migrants in rural areas compared to urban settings. Rural Australians were overall somewhat less distressed and less likely to have reported a pre-existing mental health condition (not an uncommon finding for rural regions (Smith et al., 2008)). They were also more likely to be married and to be satisfied with their local community.

Table 1
Distribution of extreme conditions and dryness patterning in Australian population sample (N=5012).

Drought pattern	Description of conditions	% Population exposed
Zero to moderate	Drought was either not present or not extreme.	47.26
Very dry	Drought period(s) were intensely dry (top 10% of sum-method drought periods).	12.30
Recent long period	An unbroken period of relative dryness lasted for at least 12 straight months in the last 24 months of the 'Big Dry'.	30.10
Constant	Cumulatively experienced 20–32 months in drought that met both the sum and count method definition (top 10% of months in drought).	3.41
Constant and recent long period	Both 'Recent long period' and 'Constant' drought conditions are present.	6.94

Table 2
Characteristics of rural and urban respondents in Wave 7 of the Household Income and Labour Dynamics in Australia Survey who have lived in the same location for at least seven years.

Variables	Proportion[a] (%)	Rural (N=919)	Urban (N=4093)
Distress (K10)	*Mean*	15.08 (0.28)	15.61 (0.14)
Age (years)	15–25 years	13.34 (0.02)	16.79 (0.01)
	26–39 years	10.20 (0.01)	10.43 (0.01)
	40–55 years	33.04 (0.02)	31.38 (0.01)
	56–65 years	22.73 (0.02)	18.50 (0.01)
	66–79 years	17.97 (0.02)	17.15 (0.01)
	80+ years %	2.72 (0.01)	5.76 (0.01)
Gender	Male	51.62 (0.01)	47.28 (0.01)
Ethnicity	Indigenous/Torres Strait Islander Australian	1.54 (0.01)	1.32 (0.004)
	English-speaking immigrant	9.31 (0.01)	9.77 (0.01)
	Non-English-speaking immigrant	5.17 (0.01)	17.08 (0.01)
Left school at 15 years or younger		42.08 (0.02)	36.22 (0.01)
Education	Year 11 or less	45.20 (0.02)	37.25 (0.01)
	Year 12	14.96 (0.01)	14.55 (0.01)
	Certificate or diploma	30.49 (0.02)	28.77 (0.01)
	Tertiary degree	3.58 (0.01)	11.11 (0.01)
	Higher degree	5.77 (0.01)	8.32 (0.01)
Married		56.61	66.05
Gross household income (AUD$)	nil-$31,199	23.54 (0.03)	19.32 (0.01)
	$31,200–$51,999	22.47 (0.03)	16.74 (0.01)
	$52,000–$77,999	19.13 (0.03)	23.43 (0.01)
	$78,000–$114, 399	22.11 (0.03)	21.86 (0.01)
	$114,400+	12.76 (0.03)	18.65 (0.01)
Pensions/allowances > 30% of income		31.57 (0.04)	29.68 (0.01)
Health condition		31.12 (0.02)	30.63 (0.01)
Mental health condition		2.25 (0.005)	3.69 (0.004)
Farm-related household		15.58 (0.03)	0.82 (0.002)
Satisfaction with local community (average 0–10)	*Mean*	7.46 (0.10)	6.92 (0.06)
Months in sum drought	*Mean*	14.15	14.54
Months in count drought	*Mean*	19.21	20.86
Months in drought (both methods)	*Mean*	13.85	14.48
Dryness intensity measure	*Mean*	5.58	5.80
No. months in recent count index dryness	*Mean*	4.93	7.15
Dryness pattern	Zero to moderate	65.54	44.31
	Very dry	5.49	13.44
	Recent long period only	18.08	32.10
	Constant drought only	5.83	3.00
	Constant drought and recent long period	5.63	7.15

Note 1: Although equivalised income was included in analysis, for clarity and descriptive purposes income is displayed here in terms of distribution across five different income brackets.
Note 2: Robust standard errors generated by a clustered and weighted survey design are included in brackets for the person-related variables.
[a] Unless otherwise specified.

In terms of their environment, urban dwellers less likely to live in conditions classified as zero-to-moderate drought and more often experienced intensely dry periods or a recent long period of unbroken dryness. However, urban and rural areas were similar in the total number of months spent in drought (see also Table C, in the online supplementary materials).

3.3. Dryness patterns and mental health

In a multiple regression model, adjusted for confounders, the interaction between locality and drought type was statistically significant (reference category: zero-to-moderate: t=2.22, p=0.03). Mean distress scores for respondents in the five drought pattern categories split by locality are shown in Fig. 1. Fig. 1 also includes two shaded bands showing the moderate distress and low distress diagnostic bands of the Kessler-10 measure. (Note that the confidence intervals displayed should not be used to compare across drought categories because confidence intervals for mean values are calculated using standard errors rather than the square-rooted standard errors used for t-statistic confidence intervals that are appropriate when comparing across means.) For rural participants, only those in the 'constant' and 'constant and recent long' drought categories had a mean distress confidence interval that crossed into the moderate distress band. In fact, rural participants in the 'constant and recent long' category had a mean distress confidence interval that was above the diagnostic low distress band of the Kessler-10 distress measure. These mean scores suggested that

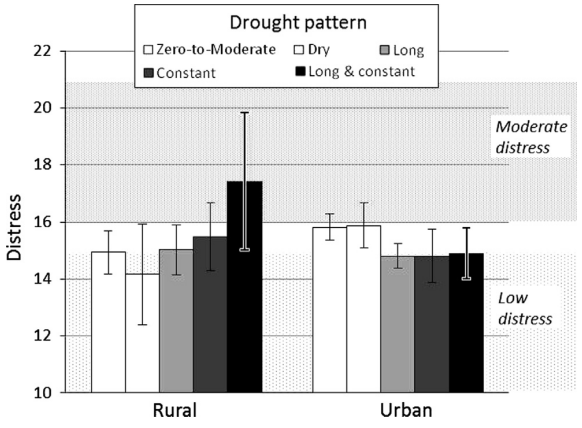


Fig. 1. Mean distress for rural and urban participants experiencing different dryness patterns. *Note*: Confidence intervals should be used to assess distress levels in terms of the Kessler 10 distress classifications 'low' and 'moderate' rather than to compare distress across categories.

Table 3
Distress predictors for rural and urban respondents with zero-to-moderate drought as the reference category.

	Rural			Urban		
	Regression coefficient	95% CI		Regression coefficient	95% CI	
Dryness pattern (Ref.: zero-to-moderate)						
Dry	−0.29	−2.11	1.52	−1.94 e−03	−0.67	0.66
Long	−0.51	−1.48	0.47	−0.17	−0.68	0.35
Constant	0.31	−0.73	1.35	−0.53	−1.22	0.15
Long & constant	2.49	0.59	4.39	−0.81	−1.70	0.09
Age (Ref.: 15–25 years)						
26–39 years	−1.18	−3.26	0.89	0.06	−1.00	1.13
40–55 years	−0.60	−2.14	0.94	−0.85	−1.77	0.07
56–65 years	−0.98	−3.10	1.14	−1.83	−2.84	−0.82
66–79 years	−3.02	−4.62	−1.42	−2.40	−3.33	−1.48
80+years %	−3.63	−5.94	−1.32	−3.44	−4.53	−2.36
Female	0.49	−0.14	1.11	0.79	0.38	1.21
Married	−1.07	−2.37	0.22	−1.06	−1.59	−0.52
Equivalised household income	−4.58 e−06	1.69 e−05	7.70 e−06	1.08 e−05	1. 08 e−05	−3.68 e−06
Non-English speaking background	1.17	−0.25	2.59	2.45	1.66	3.24
Australian Aboriginal	−1.89	−3.12	−0.67	−1.77	−3.67	0.13
Education (Ref.: 11 years or less)						
Year 12	−0.40	−1.54	0.75	0.22	−0.69	1.13
Certificate or diploma	−0.14	−1.08	0.80	−0.61	−1.17	−0.05
Tertiary degree	−0.62	−1.77	0.52	−0.87	−1.57	−0.16
Higher degree	0.30	−1.20	1.80	−0.85	−1.65	−0.05
Health condition	2.63	1.58	3.68	2.44	1.86	3.03
Mental health condition	6.24	3.15	9.32	6.81	4.87	8.75
Satisfaction with local community	−0.89	−1.23	−0.55	−0.45	−0.57	−0.34
Constant	22.62	19.62	25.63	19.64	18.56	20.72
R^2	0.25			0.21		

Note: Multiple regression analyses were adjusted for survey clustering at the collection district level and HILDA Survey sample weights.

rural respondents were more likely to have increased distress if they experienced an extreme number of months in drought, particularly if they also experienced a recent long period of unbroken dryness. In contrast, in the urban population, there was a trend for *less* distress in the 'long', 'constant-only' and 'constant and recent long' drought categories, with these means having confidence intervals below the moderate distress band.

A multiple linear regression analysis conducted within the rural sub-population indicated that people in the 'constant and recent long' drought category tended to experience more distress than those in the 'zero-to-moderate' drought category (Table 3). Based on 95% confidence intervals, the rate of change for the conditional mean of distress with respect to 'constant and recent long' drought was estimated to be an increase of 6.22% in distress compared with other participants (95% Confidence Interval: 1.46–10.98). This was substantial: nearly half the size of the rate of change for the conditional mean of distress with respect to having a pre-established mental health condition; estimated to be an increase of 15.59% more distress (95% CI: 7.87–23.31).

In contrast, a multiple linear regression analysis conducted within the urban sub-population indicated that the trend for people to have less distress in 'constant and recent long' drought could not quite be accepted with 95% confidence because the confidence interval for the estimated effect crossed zero (Table 3). Amongst the confounders, there were also non-trivial estimates for the association between worse mental health and several confounders: being of Indigenous Australian Ethnicity, being younger, having a health condition and being dissatisfied with one's local community. Although interaction terms were not tested as part of the planned analyses, readers should note that the higher estimates in the urban sample for the association between mental health and income, education and relationship status is consistent with the different lifestyle demands of rural and urban localities in Australia.

Table 4
Summary table for the relationship between distress and rural/urban locality within the five dryness pattern categories.

Dryness pattern	Regression coefficient	95% CI	
Zero-to-moderate	−0.01	−0.74	0.72
Dry	−0.22	−1.76	1.32
Long	−0.36	−1.17	0.44
Constant	0.77	−0.24	1.78
Long and constant	2.69	0.43	4.95

Note: Within each dryness pattern category, distress was regressed on locality (rural=1) using a multiple regression analysis with survey clustering and weighting, adjusting for all covariates.

Comparisons across locality, also using multiple regression analyses adjusted for confounders (Table 4), indicated that rural people experiencing the 'constant and recent long' drought pattern tended to be more distressed than were urban people living in the same conditions. The rate of change for the conditional mean of distress with respect to locality for people living in 'constant and recent long' drought was estimated to be a percentage change of 6.73% (95% CI: 1.08–12.38%). The similar but much weaker trend for 'constant' drought to be related to worse mental health in rural but not urban localities could not be accepted with 95% confidence because the confidence interval for the estimated effect crossed zero. Similarly, mean distress in 'zero-to-moderate' drought tended to be lower in rural localities, but overlapping confidence intervals indicated that this difference was not necessarily meaningful.

4. Discussion

We hypothesised that drought would become an exposure for mental health when drought was examined across a long period of

time, taking into account patterns of different drought characteristics and underlying capacity for resilience. Our findings most strongly showed that an extreme cumulative number of months in drought period conditions (20–32 months) over the seven-year 'Big Dry', culminating in a recent period of unbroken dryness lasting a year or more, was associated with increased distress in rural areas (i.e., the 'constant and recent long' drought pattern). The importance of locality was consistent with our second hypothesis, that drought would have an amplified effect in rural areas where residents would be more sensitive to drought conditions, while no consistent effects for drought were found in urban areas.

Our analysis benefited from examining relative dryness over a longer time period than has been examined in past quantitative studies examining drought and mental health. However, our more novel contribution was to show that the pattern of extreme dryness occurring within a drought itself matters for mental health. Hence, an association between drought and mental health was found for rural people experiencing the 'constant and recent long' drought pattern but not the 'constant' or 'long' drought pattern, despite each of these categories involving a similar number of total months in drought. On a clinical level, rural people living with 'constant and recent long' drought tended to experience a subclinical level of moderate distress, indicating that this subpopulation had an increased risk of developing a mental health problem requiring treatment in the future.

We had conservatively hypothesised that rural locality would amplify people's responses to different drought conditions; the differences were even more striking than we had anticipated. Rather than having a similar but less intense pattern of responses, urban dwellers showed a mean-level trend to do *better* in response to spending more time in the kind of drought pattern that tended to elevate the distress of rural people. Confidence intervals indicated that this trend in the distress levels of urban participants was highly variable, and both the presence and variability of this trend highlights the very different practical and psychological significance that drought has for urban versus rural society. Drought in the country means failing crops and starving livestock, with obvious and potentially disastrous flow-on effects for farm enterprises and rural communities. In contrast, while city-dwellers are not necessarily insensitive to the impact of drought, the endless sunny days may sometimes simply mean more opportunities to get together out-of-doors with friends.

In the introduction to this study, we suggested that drought would act as a mental health exposure when it challenged people's resilience. The rural communities in our study, adapted though they are to Australia's harsh conditions, appeared to have particular difficulty managing life under 'constant and recent long' drought conditions. The noticeable similarity between distress levels in all other comparisons suggests that it is not how dry the weather gets during periods of relative dryness that matters but how long the weather stays dry. Taken separately, a drought period lasting over a year or multiple drought periods cumulatively adding to many months in drought were not associated in our study with discernible population-level mental health harm. However, our findings suggest that people may begin to doubt their capacity to outlast a drought that initially has multiple drought periods and then settles into unbroken dryness (which may continue indefinitely).

On a positive note, all of the city people and most of the rural people in our sample appeared to be able to manage the pressures of the 'Big Dry' without evident mental health impacts. However, in the context of climate change, dealing with an increasingly drought stricken 'new normal' is a challenge that rural societies and policy-makers must face the world over. Given likely future warming and drying scenarios, the limits of people's resilience in certain settings must be considered and addressed as an emerging priority in population health.

4.1. Limitations and future directions

This study is a first step towards quantifying how wellbeing will be affected by projected increases in the frequency or intensity of drought. Our methodology needs to be replicated using data from many different countries and larger samples. It could also be expanded using person-tracking across space to quantify the weather experience of migratory people, since their socio-economic and personal characteristics may differ systematically from people who stay in the same place for very long periods. In addition, while data are often not available, more comprehensive quantification of weather conditions (e.g., temperature, humidity, wind speed and direction and soil moisture) would provide a richer and potentially more robust quantification of drought as a mental health exposure. In particular, though we have appropriately examined the measurement of *relative* dryness, it will be necessary to simultaneously investigate the potential for *absolute* measures to explain variance in mental health outcomes, or to control for absolute conditions in considering relative exposures.

Finally, this paper does not address the mechanisms by which drought has its impact. Low rainfall creates the risk that people will not have sufficient water to meet their needs but the capacity, water level and accessibility of local reservoirs and other fresh water sources, such as rivers, will moderate the level of this risk. In addition, while rural communities are much more likely to be vulnerable to weather conditions because of their close connection to the agricultural industry and distance from services, there is much unexamined variation in the strength and nature of these relationships. For example, agricultural pursuits have different water requirements (e.g. dryland versus irrigated farming) and the timing of rainfall can be as important as the amount of rain. Finally, the way that people and communities deal with scarce resources will be affected by factors like adequate mental and physical health services, social capital and the ability and willingness to be innovative in finding new ways to prosper under changed circumstances (OBrien et al., 2012; Schirmer et al., 2013). Understanding the social and economic ways in which these kinds of factors moderate and mediate the impact of drought is a complex project, probably best explored by longitudinally examining rural and urban communities over long periods of time using customised data collections.

4.2. Conclusions

The study of drought and mental health is an emerging field and, to our knowledge, this is the first paper that identifies an association between explicit, modelled measures of drought and levels of distress in the resident population. We took the novel approach of examining cumulative rainfall conditions over a relevant and long period (the critical seven years of Australia's 'Big Dry') and identifying different patterns of extreme drought. Using robust and validated relative dryness indices as a starting point, we calculated and examined many different characteristics of drought. We found that an extreme number of months in relative dryness culminating in a long period of unbroken drought was associated with an accumulated mental health exposure for people living in rural areas.

Acknowledgment

The research reported in this article was partly supported by a grant from the Australian National Health and Medical Research Council (NHMRC 585482) paid to the Australian National University, which funded a Research Fellowship position for Léan OBrien. The NHMRC is a Federal Australian Government institution. There are no other interests to declare.

Appendix A. Supplementary materials

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.envres.2014.03.014.

References

ABS, 2011. Australian Statistical Geography Standard (ASGS): Volume 4 – Significant Urban Areas, Urban Centres and Localities, Section of State. Australian Bureau of Statistics, Canberra.
Berry, H.L., et al., 2010. Climate change and mental health: a causal pathways framework. Int. J. Public Health 52, 123–132.
Berry, H.L., et al., 2008. Social Policy Research Paper: Contemporary Australian Archetypes – Different People, Different Needs, vol. 32. Commonwealth of Australia, Canberra.
Connor, L., Albrecht, G., Higginbotham, N., Freeman, S., Smith, W., 2004. Environmental change and human health in upper Hunter communities of New South Wales, Australia. EcoHealth 1, SU47–SU58.
Coombs, T., 2005. Australian Mental Health Outcomes and Classification Network Kessler – 10 Training Manual. NSW Institute of Psychiatry, Sydney.
Cuijpers, P., Smit, F., 2004. Subthreshold depression as a risk indicator for major depressive disorder: a systematic review of prospective studies. Acta Psychiatr. Scand. 109, 325–331.
Dai, A., 2010. Drought under global warming: a review. WIREs Clim. Change 2, 45–65.
Edwards, B., et al., 2008. Social and Economic Impacts of Drought on Farm Families and Rural Communities: Submission to the Productivity Commission's Inquiry into Government Drought Support. Australian Institute of Family Studies, Melbourne.
Enders, C.K., 2001. The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. Psychol. Methods 6, 352–370.
Hanigan, I.C., 2012. Monthly drought data for Australia 1890–2008 using the Hutchinson Drought Index. Australian Data Archive, http://dx.doi.org/10.4225/13/50BBFD7E6727A.
Hanigan, I.C., 2012. Suicide and drought in New South Wales, Australia, 1970–2007. Proc. Natl. Acad. Sci. 109, 13950–13955.
Hennessy, K., et al., 2008. An assessment of the impact of climate change on the nature and frequency of exceptional climatic events.
Hobfoll, S.E., 2002. Social and psychological resources and adaptation. Rev. Gen. Psychol. 6, 307–324.
Hossain, D., et al., 2008. The mental health of landholders in Southern Queensland: issues and support. Aust. J. Rural Health 16, 343–348.
Kelly, B.J., et al., 2011. Determinants of mental health and well-being within rural and remote communities. Soc. Psychiatry Psychiatr. Epidemiol. 46, 1331–1342.
Kessler, R.C., et al., 2002. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol. Med. 32, 959–976.
Nicholls, N., et al., 2006. Inter-annual rainfall variations and suicide in New South Wales, Australia, 1964–2001. Int. J. Biometeorol. 50, 139–143.
OBrien, L.V., et al., 2012. The structure of psychological life satisfaction: insights from farmers and a general community sample in Australia. BMC Public Health 12, 976–987.
Pearlin, L.I., 1999. The stress process revisited: Reflections on concepts and their interrelationships. In: Aneshensel, C.S., Phelan, J.C. (Eds.), Handbook of the Sociology of Mental Health. Springer, New York, pp. 395–416.
Rigby, C., et al., 2011. If the land's sick, we're sick: the impact of prolonged drought on the social and emotional well-being of Aboriginal communities in rural New South Wales. Aust. J. Rural Health 19, 249–254.
Sartore, G., et al., 2005. Wisdom from the drought: recommendations from a consultative conference. Aust. J. Rural Health 13, 315–320.
Schirmer, J., et al., 2013. Healthier land, healthier farmers: considering the potential of natural resource management as a place-focused farmer health intervention. Health Place 24, 97–109.
Slade, T., et al., 2011. Kessler psychological distress scale: normative data from the 2007 Australian National Survey of Mental Health and Wellbeing. Aust. N. Z. J. Psychiatry 45, 308–316.
Smith, D., et al., 1992. Climatic and Agricultural Drought: Payments and Policy. Centre for Resource and Environmental Studies, Australian National University, Canberra, ACT.
Smith, K.B., et al., 2008. Addressing the health disadvantage of rural populations: how does epidemiological evidence inform rural health policies and research? Aust. J. Rural Health 16, 56–66.
Stain, H.J., et al., 2011. The psychological impact of chronic environmental adversity: responding to prolonged drought. Soc. Sci. Med. 73, 1593–1599.
Stanke, C., et al., 2013. Health Effects of Drought: a Systematic Review of the Evidence PLOS Currents Disasters. June 5 edition 1.
Verdon-Kidd, D.C., Kiem, A.S., 2009. Nature and causes of protracted droughts in southeast Australia: comparison between the Federation, WWII, and Big Dry droughts. Geophys. Res. Lett. 36, 6.
WHO, 2005. Promoting Mental Health. World Health Organisation, Geneva.
Wooden, M., et al., 2002. The Household, Income and Labour Dynamics in Australia (HILDA) Survey: wave 1. Aust. Econ. Rev. 35, 339–348.

3.7 Paper 3: The lack of effect on women's depression of drought in Australia

Title: Powers, J. R., Dobson, A. J., Berry, H. L., Graves, A. M., Hanigan, I. C., and Loxton, D. (2015). Lack of association between drought and mental health in a cohort of 45-61 year old rural Australian women. Australian and New Zealand Journal of Public Health. 39(6):518-23.

Stage of publication: Published peer-reviewed paper

Journal Impact Factor: 2

My contribution: I contributed the Hutchinson Drought Index exposure data that were used by this paper. I assisted the authors to describe my original data. I wrote a software tool to extract the drought index values for each pixel overlayed by the geo-coded health outcome data. This was implemented using the online restricted GIS database server which I have described in section 3.2 of this thesis (Conference presentation and working paper: 'Open software - restricted data'). I also contributed to the drafting and critical review of the manuscript.

Approvals required for inclusion of Paper 3 in this thesis:

Name	Notes	Signature/date
Ms. Jennifer R. Powers		2015-08-17
Prof. Annette J. Dobson		2015-09-29
Prof. Helen L. Berry		2015-12-08
Ms. Anna M. Graves		2015-10-06
Mr. Ivan Hanigan		2015-09-28
Prof. Deborah Loxton		2015-10-06
Copyright	License Number: 3723571126268 License date: Oct 07, 2015 Licensed Content Publisher: John Wiley and Sons	2015-10-07
Ethics Approval		

MENTAL HEALTH

Lack of association between drought and mental health in a cohort of 45–61 year old rural Australian women

Jennifer R. Powers,[1] Annette J. Dobson,[2] Helen L. Berry,[3] Anna M. Graves,[1] Ivan C. Hanigan,[4] Deborah Loxton[1]

Abstract

Objective: To evaluate the impact of drought on the mental health of rural Australian women and those in vulnerable sub-populations: women who were more isolated, poorer and less educated; and women who had histories of chronic disease or poor mental health.

Methods: Surveys were mailed in 1996, 1998, 2001, 2004 and 2008 to 6,664 women born between 1946 and1951 who were participating in the Australian Longitudinal Study on Women's Health. The surveys included the Mental Health Index of the Medical Outcomes Study Short-Form 36 (MHI). Drought was assessed by linking the latitude and longitude of women's place of residence at each survey to the Hutchinson Drought Index. Associations between MHI and drought were assessed using linear mixed-models.

Results: While 31% of the women experienced drought in 1998 and 50% experienced drought in 2007; experience of droughts was less common in the other years. Although drought varied from survey year to survey year, mental health did not vary with drought conditions for rural women or vulnerable sub-populations.

Conclusions: These findings are contrary to the long-held assumption that droughts increase mental health problems in Australia.

Implications: While similar results may not be true for men, empirical evidence (rather than assumptions) is required on associations between drought and mental health.

Key words: mental health, drought, women, vulnerable

Mental health is an important public health issue.[1] About one in four people experience at least one mental health problem in any given year in the United States[2] and in England.[3] Australian mental health surveys indicate the prevalence of mental health problems has remained at 20% for a decade.[4,5] Any factor that is likely to adversely affect mental health is a major concern.

Drought has been hypothesised to lead to increased mental health problems indirectly due to its damaging effects on physical health and, especially, on the economic and social base of communities; and directly through feelings of chronic loss and failure.[6-8] This hypothesis is supported by qualitative findings of poorer mental health associated with drought[9,10] and destruction of the landscape.[11] Quantitative evidence on the effect of drought on mental health is scant[12] and inconclusive.[7,13-16] In cross-sectional drought studies, no effect was found on the mental health of 2,462 rural adults,[14] or of 6,584 rural women,[16] although increased anxiety and emotional distress was found in a small Brazilian study of 204 adults.[7] In New South Wales, Australia, suicide rates were found to be associated with lower rainfall (1964–2001),[15] and were higher in 10–49 year old rural males, but not rural females, exposed to drought between 1970 and 2007.[13] Stronger evidence of the effect of drought on mental health may be provided by longitudinal studies.

Evidence is mounting of more intense droughts,[17] particularly between 25 and 40 degrees latitude,[18] with considerably decreased rainfall resulting in more intense droughts in the mid-latitudes of Australia, the Mediterranean, Central America and Mexico, central North America and southern Africa.[19,20] Furthermore, drought and other adverse weather events are expected to differentially affect vulnerable sub-populations.[6,21] A number of factors contribute to vulnerability, including geographic isolation, lower socioeconomic status and poorer health, and these may all result in increased risk of adverse health effects as a result of climate change. In addition, mental health problems are more common among socially disadvantaged people,[22] and so adverse weather events are expected to have a greater effect on the mental health of vulnerable subpopulations.[23]

Longitudinal data are needed to strengthen the evidence base for policy.[24] By using longitudinal data, the effects of drought on mental health can be investigated in the same individuals, effectively using the individuals

1. Research Centre for Gender, Health and Ageing, University of Newcastle, New South Wales
2. School of Population Health, University of Queensland
3. Faculty of Health, University of Canberra, Australian Capital Territory
4. National Centre for Epidemiology and Population Health, Australian National University, Australian Capital Territory
Correspondence to: Ms Jennifer Powers, Research Centre for Gender, Health and Ageing, University of Newcastle, Callaghan, NSW 2308; e-mail: jenny.powers@newcastle.edu.au
Submitted: July 2014; Revision requested: October 2014; Accepted: January 2015
The authors have stated they have no conflict of interest.
Aust NZ J Public Health. 2015; Online; doi: 10.1111/1753-6405.12369

© 2015 Public Health Association of Australia

as their own controls. This paper uses 12 years of longitudinal data to: 1) investigate whether drought is associated with poorer mental health among women living in rural Australia; and 2) determine whether drought is more likely to affect the mental health of women who are more geographically isolated, poorer and less educated or have histories of chronic disease or poor mental health.

Methods

Participants

Potential participants were invited to join the Australian Longitudinal Study on Women's Health in 1996. Women were randomly selected from the national health insurance database, Medicare Australia, which includes all Australian citizens and permanent residents. To ensure that women living in rural and remote areas were adequately represented, they were sampled at twice the rate of women in urban areas.

The first mailed survey was completed by 13,715 women aged 45 to 50 years (the 1946–1951 cohort), with an estimated response of 53% to 56%.[25] Respondents were broadly representative of women of the same age in the Australian population, with some over-representation of Australian-born and educated women.[25] Women were resurveyed in 1998 (47–52 years), 2001 (50 to 55 years), 2004 (53–58 years) and 2007 (56–61 years). Contact was maintained by newsletters, postal mail, telephone and email. Non-respondents were tracked using multiple methods[26] and vital status of non-respondents was established by annual linkage to the National Death Index.[27] Further details of the study are available online.[25] Ethics approvals were obtained from the University of Newcastle (H0760795) and University of Queensland (2004000224). Written informed consent for participation in the study was obtained from participants.

The Accessibility/Remoteness Index of Australia was used to classify the women's area of residence: major cities; inner regional; outer regional; or remote or very remote.[28] Women were included in the present analyses if they were living outside major cities in 1996 and between 25 and 40 degrees south of the equator, as these are the latitudes where drought is commonly experienced in Australia.[29] In 1996, 6,664 women (49% of the total sample of 13,715) met these inclusion criteria.

Measures

Unless indicated otherwise, all measures were taken at all five surveys.

Mental health: Mental health was measured using the Mental Health Index of the well-validated Medical Outcomes Study Short Form 36 (SF36).[30] The Mental Health Index was used for these analyses because it is a valid and reliable measure[30,31] and has been shown to distinguish between groups differing in the presence and severity of psychiatric disorders.[32] Furthermore, it is responsive to change over time.[33-35] The Mental Health Index contains five items measuring: nervousness; low mood; feeling down; feeling calm and peaceful; and being happy. Scores range from 0 to 100, with higher scores indicating better mental health. In these analyses the Mental Health Index was used as both a continuous and categorical measure. A score of less than 53 for the Mental Health Index is indicative of depression[30,36] and was used to define a subgroup of women with poor mental health. Normative data from the 1995 Australian National Health Survey show that a four-point lower mean score for the Mental Health Index is associated with the presence of a serious physical condition;[37] this information provides a benchmark for assessing the clinical importance of differences between groups of women.

Drought: Many measures exist for defining climatic drought, including the Palmer Drought Severity Index[38] and annual rainfall deciles,[39] and a modified version of these two indices – the Hutchinson Drought Index.[40] The Hutchinson Drought Index counts consecutive months of lower-than-median rainfall based on percentiles of rainfall recorded at each location.[13,40] A comparison of NSW agricultural drought declarations and the Hutchinson Drought Index, published in 1992, showed good agreement when the Hutchinson Drought Index score indicated at least ten months of abnormally dry conditions. Another advantage of the Hutchinson Drought Index is that it allows valid comparisons of drought across different regions and different years.[40] The Hutchinson Drought Index was added to participant data based on the latitude and longitude of each woman's place of residence[41] at the time she completed each of the five surveys. Women were defined as living with drought if they had experienced at least ten months of abnormally dry conditions. Over the 12 years of this study, women could experience drought at one or more of the surveys and no drought at other surveys, thus acting as their own control in longitudinal data analysis.

Explanatory factors: Several factors might mediate the impact of drought on mental health. Demographic factors included age and educational attainment in 1996, and area of residence, relocation to a major city, ability to manage on available income and relationship status at each of the five surveys. Highest level of education attained in 1996 was categorised as: 10 or less years of school; 11–12 years of school, apprenticeship or trade; certificate or diploma; or university. Ability to manage on available income was classed as: always difficult; difficult some of the time; not too bad; and easy. Relationship status was classified as: married or living in a de facto relationship; separated or divorced; widowed; or single.

Health-related factors that may mediate the impact of drought on mental health included tobacco and alcohol intake measured at every survey. Smoking categories were: never smoker; ex-smoker; current smoker of fewer than 10 cigarettes a day; or current smoker of 10 or more cigarettes a day. Alcohol intake was based on reported frequency (drink rarely; less than once a week; 1 or 2 days a week; 3 or 4 days a week; 5 or 6 days a week; or every day) and quantity of alcohol usually consumed per day (1 or 2 drinks; 3 or 4 drinks; 5 to 8 drinks; or 9 or more drinks) measured at all surveys except 2001. At the 2001 survey, alcohol intake (grams/day) was measured using the Cancer Council of Victoria food frequency questionnaire. Three categories were defined: non-drinker; moderate drinker (up to 14 drinks per week); and heavy drinker (15 or more drinks per week). Women were defined as having chronic disease if they reported in 1996 being diagnosed with any of the following: diabetes; heart disease; hypertension; stroke; bronchitis; asthma; or osteoporosis. In addition, experience of partner violence (ever in a violent relationship) may affect the impact of drought on mental health. Whether a woman had ever experienced partner violence was available for every survey.

Statistical analysis

All analyses were conducted in SAS version 9.3 (SAS Institute, Cary, NC). Associations between categorical explanatory factors and drought in 1996 were tested using chi-squared. The general linear model in SAS was used to test associations between the explanatory factors and continuous Mental Health Index scores. Explanatory factors were considered potential intermediary factors if

© 2015 Public Health Association of Australia

they were significantly associated with either drought or with Mental Health Index scores. Linear mixed models (the MIXED procedure in SAS) were used to analyse the longitudinal data, taking into account the correlated data within individuals. Up to five Mental Health Index scores were available per participant between 1996 and 2007. Mixed-models were fitted with random effects (intercept) for individual women, with all other effects fixed. Robust standard errors were estimated. The outcome measure was continuous Mental Health Index scores and the main explanatory variable was drought. Models were adjusted for demographic and health-related factors and experience of partner violence at each survey. All models included survey year (fitted as a categorical variable with 1996 as the reference year). Interaction terms for time (survey year) and drought were fitted to models where both main effects were statistically significant.

Few data were missing for individual variables at any survey, and 95% of the women answered at least two of the five surveys. Missing data, mainly due to loss to follow-up, were addressed by the use of the MIXED procedure[42] and the following sensitivity analyses. Complete case analyses were conducted for all eligible women. Separate models were also fitted for potentially vulnerable sub-groups of women identified as isolated, poorer, less educated, and women with histories of chronic disease or poorer mental health based on their characteristics measured in 1996. Isolated women were defined as those living in outer regional or remote areas, the most geographically isolated and furthest from services. Poorer women were those who said they always had difficulty managing on available income. Less-educated women were those who reported their highest educational level was up to 10 years of school. Women with a history of chronic disease were those with at least one chronic condition, and women with poor mental health reported a score of less than 53 for the Mental Health Index in 1996. For all sub-groups, the sample sizes were adequate to detect a clinically significant difference of four points or more in the Mental Health Index at 5% significance level with 80% power.

Results

Of the 6,664 women included in the current study, loss to follow-up between 1996 and 2007 was due to death or incapacity (n=177), withdrawal (n=411) or lack of contact (n=320) during the 2007 survey. Eighty-two per cent of the women (n=5,441) completed four or all five surveys. Completion and loss to follow-up were not associated with drought. While drought was not common in 1996, 2001 and 2004, one-third of the women experienced drought in 1998 and half the women experienced drought in 2007 (Figure 1). Thirty-five per cent of women did not experience drought at any survey; 43% experienced drought at one survey; 20% at two surveys; 2% at three surveys; two women at four surveys; and none experienced drought at all five surveys. Being in drought in 1996 was associated with area

Figure 1: Percentages of women living in drought and their mean Mental Health Index at each survey.

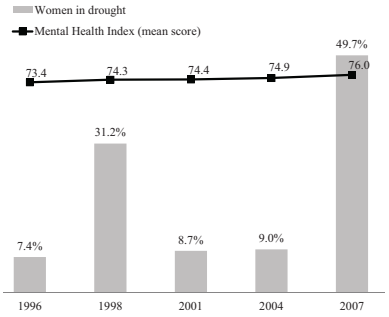


Table 1: Characteristics of 6,664 Australian women by drought[a] and Mental Health Index in 1996.

	N	Not in drought N=6172 %	Drought N=492 %	Mental Health Index mean (95% CI)
Demographic factors				
Area of residence		$p<0.01$		$p=0.05$
Inner regional	4,492	68.5	53.5	73.1 (72.6, 73.6)
Outer regional	1,871	27.9	29.9	73.7 (72.9, 74.5)
Remote or very remote	301	3.6	16.7	75.5 (73.5, 77.5)
Relocation to a major city between 1996 and 2007		$p=0.81$		$p=0.19$
Yes	690	10.9	10.5	73.0 (71.6, 74.3)
No	5,666	89.1	89.5	73.9 (73.4, 74.3)
Highest educational level		$p<0.01$		$p<0.01$
10 years school or less	3,598	54.9	48.6	72.0 (71.4, 72.6)
11-12 years school or trade	1,213	17.9	23.8	74.6 (73.7, 75.6)
Certificate or diploma	1,013	15.4	14.1	76.1 (75.0, 77.2)
University	783	11.7	13.5	74.6 (73.4, 75.8)
Ability to manage on income		$p=0.07$		$p<0.01$
Always difficult	1,025	15.7	12.9	63.5 (62.4, 64.5)
Difficult some of the time	2,033	30.8	28.6	71.9 (71.1, 72.6)
Not too bad	2,718	40.9	42.6	76.5 (75.9, 77.2)
Easy	854	12.6	15.9	79.0 (77.9, 80.2)
Relationship status		$p<0.01$		$p<0.01$
Married or de facto	5,686	85.1	89.6	74.3 (73.9, 74.8)
Separated or divorced	678	10.5	5.7	66.7 (65.3, 68.0)
Widowed	123	1.8	2.9	69.4 (66.3, 72.6)
Single	168	2.6	1.8	71.4 (68.7, 74.0)
Health-related factors				
Smoking status		$p=0.08$		$p<0.01$
Never-smoker	3,481	53.8	58.4	75.2 (74.6, 75.7)
Ex-smoker	1,793	27.9	27.9	73.6 (72.7, 74.4)
Smoke less than 10 cigarettes/ day	206	3.2	2.7	71.9 (69.5, 74.3)
Smoke 10 or more cigarettes/day	953	15.0	11.0	68.3 (67.2, 69.4)
Alcohol consumption		$p<0.05$		$p<0.01$
Non-drinker	1,021	15.6	13.7	73.6 (72.6, 74.7)
Moderate drinker	5,231	78.9	83.1	73.8 (73.3, 74.3)
Heavy drinker	349	5.5	3.3	68.7 (66.8, 70.6)
Any diagnosed physical condition[b]		$p=0.25$		$p<0.01$
Yes	2,976	45.2	42.5	71.2 (70.5, 71.8)
No	3,636	54.8	57.5	75.3 (74.7, 75.9)
Ever experienced partner violence		$p<0.01$		$p<0.01$
Yes	1,054	16.2	11.6	66.3 (65.3, 67.4)
No	5,592	83.8	88.4	74.8 (74.3, 75.2)

a: Drought was defined as at least 10 months abnormally dry conditions.

b: Self-reported diagnoses of diabetes, heart disease, hypertension, stroke, bronchitis, asthma or osteoporosis.

© 2015 Public Health Association of Australia

of residence, education, relationship status, alcohol consumption and experience of partner violence, but not with relocation to a major city, difficulty managing on income, smoking or history of chronic disease (Table 1). The Mental Health Index was associated with all variables except area of residence and relocation to a major city.

A small improvement was seen in the Mental Health Index between 1996 and 2007; however, the Mental Health Index did not differ by drought, even though drought varied from survey year to survey year (Table 2, Figure 1). Women who had difficulty managing on their available income and women with poor mental health in 1996 had significantly poorer Mental Health Index (Table 3) compared with the general population of women (Table 2). The Mental Health Index of women in vulnerable sub-populations improved significantly over time and was unrelated to drought (Table 3). In adjusted models, two results deviated from this pattern of findings: less-educated women and women with poor mental health in 1996 who had experienced drought had slightly better Mental Health Index, as would be expected due to regression to the mean. Complete case analyses for women who responded to all surveys provided similar results (Tables 4 and 5).

Table 2: Mixed models for mental health for 6,664 Australian women by drought.

	Unadjusted model Estimate (SE) N=6,664	Adjusted model* Estimate (SE) N=6,538
Intercept	**73.31 (0.22)**	**74.63 (0.59)**
Survey year		
1996	reference	reference
1998	**0.56 (0.21)**	0.44 (0.23)
2001	**0.44 (0.22)**	0.24 (0.24)
2004	**0.99 (0.23)**	**0.88 (0.24)**
2007	**1.80 (0.25)**	**1.52 (0.27)**
Drought status		
Not in drought	reference	reference
Drought	0.24 (0.21)	0.38 (0.21)

* *Model adjusted for age, area of residence, education, income, marital status, smoking, alcohol consumption and experience of partner violence.*

Statistically significant estimates (at 0.05 level) are shown in bold.

Two further mental health questions were asked at the 2001, 2004 and 2007 surveys: 'Have you consulted a counsellor, psychologist or social worker for your own health in the last twelve months?' and 'In the past week, have you been feeling that life isn't worth living?' Prevalence of consultation did not differ by experience of drought or not (6.0% versus 6.7% in 2001, 8.3% versus 6.7% in 2004 and 7.2% versus 8.8% in 2007). Nor were women more likely to feel life was not worth living (6.8% versus 7.8% in 2001, 6.3% versus 5.8% in 2004 and 5.4% versus 6.4% in 2007). Furthermore, drought conditions were more severe in 2007, yet women living in drought were not more likely to consult counsellors or feel that life was not worth living.

Discussion

No association was found between drought and women's mental health, either generally, or in potentially vulnerable groups. The results were unexpected given hypotheses of adverse mental health impacts as a result of drought.[6,19,43] It is possible 45–61 year old women are adaptable and cope with the chronic nature of drought without showing a direct mental health effect. In support of this alternative explanation, previous research showed no mental health deficit attributable to living in an area sufficiently affected by climate events to be eligible for relief payments.[16] Furthermore, a large New South Wales study (1970–2007) found an increased relative risk of suicide for 10–49 year old rural males with increasing drought, but not for females or older males.[13] These findings suggest that rural women are less affected by drought than men, providing hope that identifying factors that keep women healthy could inform prevention interventions for men in drought-affected communities.

The findings did not support contentions that poor mental health is likely to be exacerbated among vulnerable groups exposed to drought in high-income countries.[19,44] Vulnerability to poorer mental health is known to vary with socioeconomic and health status, age and gender[22,45] – findings

Table 3: Mixed models for Mental Health Index for subgroups of vulnerable women by drought.

	Isolated women Estimate (SE)	Poorer women Estimate (SE)	Less educated Estimate (SE)	Chronic disease Estimate (SE)	Poor mental health Estimate (SE)
Unadjusted	N=2,172	N=1,025	N=3,598	N=2,976	N=883
Intercept	**73.83 (0.38)**	**63.34 (0.64)**	**71.91 (0.31)**	**71.07 (0.34)**	**54.49 (0.76)**
Survey year					
1996	reference	reference	reference	reference	n/a
1998	0.61 (0.35)	**1.58 (0.64)**	**0.80 (0.30)**	0.37 (0.33)	reference
2001	0.45 (0.39)	**1.97 (0.68)**	**0.65 (0.32)**	**0.78 (0.34)**	**1.99 (0.80)**
2004	**0.87 (0.40)**	**3.23 (0.72)**	**0.86 (0.34)**	**1.11 (0.35)**	**3.86 (0.81)**
2007	**1.86 (0.43)**	**3.55 (0.79)**	**1.35 (0.36)**	**1.81 (0.39)**	**4.34 (0.83)**
Drought status					
Not in drought	reference	reference	reference	reference	reference
Drought	0.40 (0.37)	0.91 (0.65)	**0.74 (0.30)**	**0.64 (0.32)**	1.40 (0.76)
Adjusted model*	N=2,132	N=998	N=3,548	N=2,913	N=790
Intercept	**73.00 (1.35)**	**65.06 (1.82)**	**74.66 (5.01)**	**71.95 (0.90)**	**54.85 (2.12)**
Survey year					
1996	reference	reference	reference	reference	n/a
1998	0.60 (0.38)	**1.46 (0.69)**	**0.74 (0.34)**	0.26 (0.36)	reference
2001	0.44 (0.44)	**1.75 (0.74)**	0.64 (0.35)	0.31 (0.37)	**2.56 (0.87)**
2004	**0.94 (0.42)**	**3.39 (0.76)**	**0.97 (0.35)**	**0.97 (0.37)**	**4.40 (0.89)**
2007	**1.64 (0.47)**	**3.26 (0.84)**	**1.24 (0.38)**	**1.54 (0.42)**	**4.24 (0.91)**
Drought status					
Not in drought	reference	reference	reference	reference	reference
Drought	0.69 (0.38)	1.22 (0.67)	**0.94 (0.31)**	0.56 (0.34)	**2.15 (0.81)**

* *Model adjusted for age, area of residence, education, income, marital status, smoking, alcohol consumption and experience of partner violence.*

Statistically significant estimates (at 0.05 level) are shown in bold.

Table 4: Mixed models for Mental Health Index for complete cases (women who answered all surveys).

	Unadjusted model Estimate (SE) N=4,465	Adjusted model* Estimate (SE) N=4,439
Intercept	**75.04 (0.25)**	**75.66 (0.68)**
Survey year		
1996	reference	reference
1998	0.04 (0.24)	0.12 (0.24)
2001	-0.05 (0.25)	-0.15 (0.26
2004	0.46 (0.26)	0.51 (0.27)
2007	**1.25 (0.27)**	**1.18 (0.29)**
Drought status		
Not in drought	reference	reference
Drought	0.30 (0.22)	0.38 (0.23)

* *Model adjusted for age, area of residence, education, income, marital status, smoking, alcohol consumption and experience of partner violence.*

Statistically significant estimates (at 0.05 level) are shown in bold.

© 2015 Public Health Association of Australia

that were demonstrated for the women in this study. However, women who lived through droughts did not have worse mental health than other women, even when they were more isolated, less educated, poorer or living with chronic conditions. In a developed country such as Australia, access to resources such as information and technology, as well as the potential for financial assistance and social networks, may work to decrease vulnerability[46] and thus alleviate any mental health effects of drought. Evidence suggests that women have higher levels of social support and greater resilience than men;[47-50] factors that may explain women's ability to cope with drought.

This study has several strengths. It is the only longitudinal prospective study of the effects of drought on mental health. The measure of drought, calculated using a century of rainfall data validated against drought declarations, allowed for objective identification of drought over the period of this study. Furthermore, this study of the mid-latitudes of Australia has relevance to other parts of the world (the Mediterranean, Central America, Mexico, central North America and southern Africa) where longer-term drying is increasingly being observed.[19,20]

The data were collected prospectively over 12 years, allowing for change in mental health to be measured as drought conditions changed. The wide range of mental health scores and poorer mental health status that was evident for women in lower socioeconomic groups show that the mental health measure was responsive to differences between groups. The comprehensive nature of the surveys also allowed socioeconomic and psychosocial vulnerabilities to be taken into account.

While the large-scale, nationally representative sample of 45–61 year old women is a further strength, the findings may not apply to women of other ages. Although broadly representative of women of the same age in the Australian population, there was some over-representation of more-educated women.[25] As a result, participants would be expected to have better mental health than Australian women of the same age, which appears to be the case. Before and after adjustment for education, mental health was no worse for women living in drought and non-drought areas. Furthermore, sensitivity analysis of less-educated women revealed women living with drought had better mental health than those living in non-drought areas. These results are consistent with the findings of two smaller studies in rural and remote Australian communities.[14,51] Both studies found that the mental health of men and women was associated with individual characteristics rather than with drought.

About 20% of women were non-respondents in 2007. Not only was loss to follow-up (due to death or incapacity, withdrawal or lack of contact during the 2007 survey) the same by drought, but women were equally likely to be non-respondents across drought. While the mental health of women who were lost to follow-up may have differed by drought, complete case analyses and sub-group analyses (including women with poorer mental health) showed no association between drought and mental health.

Table 5: Mixed models for Mental Health Index for subgroups of vulnerable women (complete cases).

	Isolated women Estimate (SE)	Poorer women Estimate (SE)	Less educated Estimate (SE)	Chronic disease Estimate (SE)	Poor mental health Estimate (SE)
Unadjusted	N=1,442	N=584	N=2,262	N=1,957	N=532
Intercept	**75.40 (0.45)**	**65.69 (0.81)**	**74.02 (0.37)**	**72.96 (0.40)**	**54.65 (0.90)**
Survey year					
1996	reference	reference	reference	reference	n/a
1998	0.21 (0.39)	1.37 (0.77)	0.22 (0.35)	-0.15 (0.37)	reference
2001	0.06 (0.44)	0.97 (0.82)	0.08 (0.37)	-0.07 (0.39)	**2.30 (0.90)**
2004	0.55 (0.44)	**2.63 (0.85)**	0.31 (0.38)	0.38 (0.40)	**4.23 (0.91)**
2007	**1.37 (0.47)**	**2.82 (0.90)**	0.71 (0.40)	**1.16 (0.43)**	**4.35 (0.93)**
Drought status					
Not in drought	reference	reference	reference	reference	reference
Drought	0.48 (0.40)	1.10 (0.73)	**0.78 (0.33)**	0.56 (0.35)	**2.29 (0.85)**
Adjusted model*	N=1,433	N=583	N=2,261	N=1,946	N=526
Intercept	**74.13 (1.49)**	**67.59 (2.24)**	**68.20 (5.36)**	**72.33 (1.04)**	**54.68 (2.38)**
Survey year					
1996	reference	reference	reference	reference	n/a
1998	0.41 (0.40)	1.51 (0.77)	0.42 (0.36)	-0.09 (0.38)	reference
2001	0.09 (0.48)	1.23 (0.85)	0.21 (0.38)	-0.15 (0.41)	**2.34 (0.93)**
2004	0.76 (0.46)	**3.00 (0.88)**	0.59 (0.39)	0.50 (0.41)	**4.30 (0.97)**
2007	**1.34 (0.50)**	**2.91 (0.94)**	**0.84 (0.42)**	**1.15 (0.45)**	**3.86 (0.98)**
Drought status					
Not in drought	reference	reference	reference	reference	reference
Drought	0.62 (0.40)	1.17 (0.73)	**0.83 (0.33)**	0.59 (0.36)	**2.55 (0.88)**

** Models were adjusted for age, area of residence, education, income, marital status, smoking, alcohol consumption and experience of partner violence.*

Statistically significant estimates (at 0.05 level) are shown in bold.

Conclusions

This paper answers repeated calls for empirical studies on the effects of adverse weather events – such as drought – on mental health[52,53] and, as recommended, provides a thorough investigation of the association between mental health and drought, taking into account influences other than drought.[54,55] Over the 12 years of this longitudinal prospective study, drought was not associated with poorer mental health in mid-aged women in general, or in groups of potentially vulnerable women. The finding that, compared with men, women were at reduced relative risk of suicide during drought[13] suggests that rural women appear to be less affected by drought than rural men. As with all epidemiological studies, the findings relate to average effects and so may not hold true for individual women, whose mental health may be affected by drought.

Acknowledgements

The Australian Longitudinal Study on Women's Health, which was conceived and developed by groups of interdisciplinary researchers at the Universities of Newcastle and Queensland, is funded by the Australian Government Department of Health. We are grateful to the women who participate in the study. Researchers in the Research Centre for Gender, Health and Ageing at the University of Newcastle are members of the Hunter Medical Research Institute. Financial support to develop the drought dataset was provided by Professor AJ McMichael's 'Australia Fellowship' from the National Health and Medical Research Council.

© 2015 Public Health Association of Australia

References

1. World Health Organisation. *Depression is a Common Illness* [Internet]. Geneva (CHE): WHO; 2012 [cited 2012 Dec 3]. Available from: http://www.who.int/mediacentre/news/notes/2012/mental_health_day_20121009/en/index.html
2. Kessler RC, Wang PS. The descriptive epidemiology of commonly occurring mental disorders in the United States. *Ann Rev Public Health.* 2008;29:115-29.
3. National Centre for Social Research. *Adult Psychiatric Morbidity in England, 2007. Results of a Household Survey*. London (AUST): National Health Services Information Centre for Health and Social Care; 2009.
4. Australian Bureau of Statistics. *4326.0. - Mental Health and Wellbeing: Profile of Results, Australia 1997*. Canberra (AUST): ABS; 1998.
5. Australian Bureau of Statistics. *4326.0. - National Survey of Mental Health and Wellbeing: Summary of Results, Australia 2007*. Canberra (AUST): ABS; 2008.
6. Berry HL, Bowen K, Kjellstrom T. Climate change and mental health: A causal pathways framework. *Int J Public Health.* 2010;55:123-32.
7. Coelho AEL, Adair JG, Mocellin JSP. Psychological responses to drought in northeastern Brazil. *Interam J Psychol.* 2004;38:95-103.
8. Woodruff RE, McMichael T, Butler C, Hales S. Action on climate change: The health risks of procrastinating. *Aust N Z J Public Health.* 2006;30:567-71.
9. Alston M. Rural male suicide in Australia. *Soc Sci Med.* 2012;74:515-22.
10. Rich JL, Wright SL, Loxton D. 'Patience, hormone replacement therapy and rain!' Women, ageing and drought in Australia: Narratives from the mid-age cohort of the Australian Longitudinal Study on Women's Health. *Aust J Rural Health.* 2012;20:324-8.
11. Albrecht G, Sartore GM, Connor L, Higginbotham N, Freeman S, Kelly B, et al. Solastalgia: The distress caused by environmental change. *Australas Psychiatry.* 2007;15:S95-S8.
12. Gamble JL, Hurley BJ, Schultz PA, Jaglom WS, Krishnan N, Harris M. Climate change and older Americans: State of the science. *Environ Health Perspect* 2013;121(1):15-22.
13. Hanigan IC, Butler CD, Kokic PN, Hutchinson MF. Suicide and drought in New South Wales, Australia, 1970-2007. *Proc Natl Acad Sci.* 2012;109(35):13950–5.
14. Kelly BJ, Lewin TJ, Stain HJ, Coleman C, Fitzgerald M, Perkins D, et al. Determinants of mental health and well-being within rural and remote communities. *Soc Psychiatry Psychiatr Epidemiol.* 2011;46:1331-42.
15. Nicholls N, Butler CD, Hanigan I. Inter-annual rainfall variations and suicide in New South Wales, Australia, 1964-2001. *Int J Biometeorol* 2006;50:139-43.
16. Powers JR, Loxton D, Baker J, Rich JL, Dobson AJ. Empirical evidence suggests adverse climate events have not affected Australian women's health and well-being. *Aust N Z J Public Health.* 2012;36(5):403-98.
17. Haines A, Kovats RS, Campbell-Lendrum D, Corvalan C. Harben Lecture - Climate change and human health: Impacts, vulnerability, and mitigation. *Lancet.* 2006;367:2101-9.
18. Bi P, Parton KA. Effect of climate change on Australian rural and remote regions: What do we know and what do we need to know? *Aust J Rural Health.* 2008;16:2-4.
19. Garnaut R. *The Garnaut Climate Change Review. Final Report*. Melbourne (AUST): Cambridge University Press; 2008.
20. Intergovernmental Panel on Climate Change. Summary for Policymakers. In: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. Cambridge (UK): Cambridge University Press; 2012.
21. Balbus JM, Malina C. Identifying vulnerable subpopulations for climate change health effects in the United States. *J Occup Environ Med.* 2009;51:33-7.
22. Fryers T, Melzer D, Jenkins R. Social inequalities and the common mental disorders - A systematic review of the evidence. *Soc Psychiatry Psychiatr Epidemiol.* 2003;38:229-37.
23. Filiberto D, Wethington E, Pillemer K, Wells NM, Wysocki M, Parise JT. Older people and climate change: Vulnerability and health effects. *Generations.* 2009;33:19-25.
24. Haines A. Climate change and health: strengthening the evidence base for policy. *Am J Prev Med.* 2008;35:411-3.
25. Australian Longitudinal Study on Women's Health. *Sample and Response Rates* [Internet]. Newcastle (AUST): University of New Castle Australian Longitudinal Study on Women's Health; 2012 [cited 2012 Oct 30]. Available from: http://www.alswh.org.au/about/sample
26. Lee C, Dobson A, Brown W, Adamson L, Goldsworthy J. Tracking participants: Lessons from the Women's Health Australia Project. *Aust N Z J Public Health.* 2000;24:334-6.
27. Powers J, Ball J, Adamson L, Dobson A. Effectiveness of the National Death Index for establishing the vital status of older women in the Australian Longitudinal Study on Women's Health. *Aust N Z J Public Health.* 2000;24:526-8.
28. National Centre for Social Applications of GIS (GISCA). *ARIA and Accessibility (Accessibility/Remoteness Index of Australia)* [Internet]. Adelaide (AUST): University of Adelaide Australian Population and Migration Research Centre; 2013 [cited 2013 Jan 3]. Available from: http://www.adelaide.edu.au/apmrc/research/projects/category/aria.html
29. *Romsey Australia. Drought in Australia: A Natural Phenomenon* [Internet]. San Francisco (CA): iprimus; 2010 [cited 2013 Sep 2]. Available from: http://home.iprimus.com.au/food7/droughthistory.html
30. Ware J, Kosinski M, Keller S. *SF-36 Physical and Mental Health Summary Scales: A User's Manual*. Boston (MA): New England Medical Centre The Health Institute; 1994.
31. Berwick DM, Murphy JM, Goldman PA, Ware JE Jr, Barsky AJ, Weinstein MC. Performance of a five-item mental health screening test. *Med Care.* 1991;29:169-76.
32. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care.* 1993;31:247-63.
33. Cameron AJ, Magliano DJ, Dunstan DW, Zimmet PZ, Hesketh K, Peeters A, et al. A bi-directional relationship between obesity and health-related quality of life: Evidence from the longitudinal AusDiab study. *Int J Obes.* 2012;36:295-303.
34. Lucke JC, Brown W, Tooth L, Loxton D, Byles J, Spallek M, et al. Health across generations: Findings from the Australian Longitudinal Study on Women's Health. *Biol Res Nurs.* 2010;12(2):162-70.
35. Williams JS, Cunich M, Byles J. The impact of socioeconomic status on changes in the general and mental health of women over time: Evidence from a longitudinal study of Australian women. *Int J Equity Health.* 2013;12:25.
36. Silveira E, Taft C, Sundh V, Waern M, Palsson S, Steen B. Performance of the SF-36 Health Survey in screening for depressive and anxiety disorders in an elderly female Swedish population. *Qual Life Res.* 2005;14:1263-74.
37. Australian Bureau of Statistics. *4399.0. - National Health Survey Australia, 1995: SF-36 Population Norms*. Canberra (AUST): AGPS; 1997.
38. Quiring SM. Developing Objective Operational Definitions for Monitoring Drought. *J App Meteorol Clim.* 2009;48:1217-29.
39. Pandey RP, Dash BB, Mishra SK, Singh R. Study of indices for drought characterization in KBK districts in Orissa (India). *Hydrol Process.* 2008;22:1895-907.
40. Smith DI, Hutchinson MF, McArthur RJ. *Climatic and Agricultural Drought: Payments and Policy*. Canberra (AUST): Australian National University Centre for Resource and Environmental Studies; 1992.
41. Hanigan I. *Monthly Drought Data for Australia 1890-2008 Using the Hutchinson Drought Index*. Canberra (AUST): Australian National University Australian Data Archive; 2012.
42. Twisk J, de Boer M, de Vente W, Heymans M. Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. *J Clin Epidemiol.* 2013;66:1022-8.
43. Obrien LV, Berry HL, Coleman C, Hanigan IC. Drought as a mental health exposure. *Environ Res.* 2014;131:181-7.
44. Friel S, Bowen K, Campbell-Lendrum D, Frumkin H, McMichael AJ, Rasanathan K. Climate change, noncommunicable diseases, and development: The relationships and common policy opportunities. *Ann Rev Public Health.* 2011;32:133-47.
45. Keim ME. Building human resilience: The role of public health preparedness and response as an adaptation to climate change. *Am J Prev Med.* 2008;35:508-16.
46. Cutter SL, Boruff BJ, Shirley WL. Social vulnerability to environmental hazards. *Soc Sci Q.* 2003;84:242-61.
47. Cornwell B. Independence through social networks: Bridging potential among older women and men. *J Gerontol B Psychol Sci Soc Sci.* 2011;66:782-94.
48. Kendler KS, Myers J, Prescott CA. Sex differences in the relationship between social support and risk for major depression: A longitudinal study of opposite-sex twin pairs. *Am J Psychiatry.* 2005;162:250-6.
49. McLaughlin D, Vagenas D, Pachana NA, Begum N, Dobson A. Gender differences in social network size and satisfaction in adults in their 70s. *J Health Psychol.* 2010;15:671-9.
50. Netuveli G, Wiggins RD, Montgomery SM, Hildon Z, Blane D. Mental health and resilience at older ages: Bouncing back after adversity in the British Household Panel Survey. *J Epidemiol Community Health.* 2008;62:987-91.
51. Stain HJ, Kelly B, Carr VJ, Lewin TJ, Fitzgerald M, Fragar L. The psychological impact of chronic environmental adversity: Responding to prolonged drought. *Soc Sci Med.* 2011;73:1593-9.
52. Campbell-Lendrum D, Woodruff R. Comparative risk assessment of the burden of disease from climate change. *Environ Health Perspect.* 2006;114:1935-41.
53. Frumkin H, Hess J, Luber G, Malilay J, McGeehin M. Climate change: The public health response. *Am J Public Health.* 2008;98:435-45.
54. Huang C, Vaneckova P, Wang X, FitzGerald G, Guo Y, Tong S. Constraints and barriers to public health adaptation to climate change: A review of the literature. *Am J Prev Med.* 2011;40:183-90.
55. Patz J, Campbell-Lendrum D, Gibbs H, Woodruff R. Health impact assessment of global climate change: Expanding on comparative risk assessment approaches for policy making. *Ann Rev Public Health.* 2008;29:27-39.

© 2015 Public Health Association of Australia

Chapter 4

The effect on cardio-respiratory disease of bushfire smoke, dust and heatwaves

4.1 Paper 4: Integration of historical records of bushfire events

Title: Johnston, F., Hanigan, I., Henderson, S., Morgan, G., Portner, T., Williamson, G., and Bowman, D. (2011). Creating an integrated historical record of extreme particulate air pollution events in Australian cities from 1994 to 2007. Journal of the Air Waste Management Association, 61(4), 390-398. doi:10.3155/1047-3289.61.4.390

Stage of publication: Published peer-reviewed paper

Journal Impact Factor: 1.3

My contribution: Epidemiological studies of the health effects of biomass smoke events (such as bushfires or wood-heater smoke spikes due to inversion layers) have been hampered by the lack of availability of datasets that explicitly pertain to these sources. This study set out to address this gap in knowledge about historical smoke events in Australia. I contributed to the study design and helped develop the criteria by which air pollution events would be validated as biomass smoke or dust. I conducted all air pollution data pre-processing to clean and impute the time-series data for multiple cities using the APHEA2 protocols. I then built the events database and designed the data entry forms. I assisted Ms Portner in the validation and data entry. I summarised the validated events data and created the results summaries reported by this paper. I assisted the other authors to describe the data and helped write the paper. I gave editorial advice and critical review of descriptions of the results. I was responsible for the distribution of the associated database that was produced for this paper (dataset 3 of this thesis) as well as the software to develop the database (software package 2 of this thesis) which I host on one of my websites (https://github.com/swish-climate-impact-assessment/BiosmokeValidatedEvents).

Approvals required for inclusion of Paper 4 in this thesis:

Name	Notes	Signature/date
Dr. Fay H. Johnston		2015-08-21
Mr. Ivan Hanigan		2015-09-28
Dr. Sarah B. Henderson		2015-10-06
Prof. Geoffrey G. Morgan		2015-11-15
Ms. Talia Portner		Could not be contacted 2015-12-05
Dr. Grant J. Williamson		2015-10-06
Prof. David MJS Bowman		David Bowman 2015-10-06
Copyright	Permission granted.	2015-10-07
Ethics Approval	Not applicable	

TECHNICAL PAPER

ISSN:1047-3289 *J. Air & Waste Manage. Assoc.* **61**:390–398
DOI:10.3155/1047-3289.61.4.390
Copyright 2011 Air & Waste Management Association

Creating an Integrated Historical Record of Extreme Particulate Air Pollution Events in Australian Cities from 1994 to 2007

Fay H. Johnston
Menzies Research Institute, University of Tasmania, Tasmania, Australia

Ivan C. Hanigan
National Centre of Epidemiology and Population Health, Australian National University, Canberra, Australian Capital Territory; and School of Plant Sciences, University of Tasmania, Tasmania, Australia

Sarah B. Henderson
Menzies Research Institute, University of Tasmania, Tasmania, Australia

Geoffrey G Morgan
Northern Rivers Department of Rural Health, University of Sydney, Lismore, New South Wales, Australia; and North Coast Area Health Service, Lismore, New South Wales, Australia

Talia Portner, Grant J. Williamson, and David M.J.S. Bowman
School of Plant Sciences, University of Tasmania, Tasmania, Australia

ABSTRACT
Epidemiological studies of exposure to vegetation fire smoke are often limited by the availability of accurate exposure data. This paper describes a systematic framework for retrospectively identifying the cause of air pollution events to facilitate a long, multicenter analysis of the public health effects of vegetation fire smoke pollution in Australia. Pollution events were statistically defined as any day at or above the 95th percentile of the 24-hr average concentration of particulate matter (PM). These were identified for six cities from three distinct ecoclimatic regions of Australia. The dates of each event were then crosschecked against a range of information sources, including online newspaper archives, government and research agency records, satellite imagery, and aerosol optical thickness measures to identify the cause for the excess particulate pollution. Pollution events occurred most frequently during summer for cities in subtropical and arid regions and during winter for cities in temperate regions. A cause for high PM on 67% of days examined in the city of Sydney was found, and 94% of these could be attributed to landscape fire smoke. Results were similar for cities in other subtropical and arid locations. Identification of the cause of pollution events was much lower in colder temperate regions where fire activity is less frequent. Bushfires were the most frequent cause of extreme pollution events in cities located in subtropical and arid regions of Australia. Although identification of pollution episodes was greatly improved by the use of multiple sources of information, satellite imagery was the most useful tool for identifying bushfire smoke pollution events.

IMPLICATIONS
Landscape fire smoke is an episodic and increasingly important environmental hazard for human populations. However, the health effects attributable to landscape fire smoke remain uncertain, partially because of the challenges associated with exposure assessment. This paper describes a systematic framework for retrospectively identifying the cause of air pollution events to facilitate a long, multicenter analysis of the public health effects of vegetation fire smoke pollution in Australia.

INTRODUCTION
It is difficult to assess the potential health effects of smoke from large vegetation fires, primarily because it is challenging to assess the exposure of affected populations. The episodic nature of large events makes it impossible to know precisely when and where air quality will be affected, and exposures are often difficult to quantify retrospectively. In many cases smoke from large fires affects (1) rural areas where there are few or no air quality monitoring data to inform quantitative exposure assessment, (2) populations that are too small to support meaningful epidemiological study, or (3) large populations for durations that are too short to detect a significant effect.[1]

Despite these challenges, a growing body of literature has consistently reported that particulate matter (PM)

during episodes of vegetation fire smoke is associated with respiratory morbidity, especially related to adult asthma and chronic obstructive pulmonary disease.[2–19] This association is now well accepted.[20,21] However, the effect of PM from fire smoke on other health outcomes such as cardiovascular morbidity and mortality is uncertain. Only one[22] of a few studies[2,23,24] reports that smoke-related PM is associated with mortality, and none have reported any consistent cardiovascular effect.[2,6,9,10,14–16] The latter is surprising given the weight of evidence suggesting that smoke particles elicit toxicological effects similar to those of urban particles[25–27] and that urban PM has been causally linked with respiratory and cardiovascular morbidity and mortality.[21] One possible explanation is that the PM in vegetation fire smoke is different enough from urban PM that it carries a different public health risk. The other possible explanation is that most studies to date have been too small to detect mortality and cardiovascular effects.[1,20]

Previous researchers in this area have used a creative range of study designs and analytical approaches. Over time they have included larger study groups, covered longer periods, and developed increasingly sophisticated approaches to data analysis and exposure assessment. The authors feel that a practical way to move forward in the identification of health effects associated with vegetation smoke is to increase the size and duration of studies by collecting data from multiple centers and reaching deep into the historical records of areas habitually affected by this pollutant. Several major urban centers and smaller regional towns in the Australian states of New South Wales, Western Australia, and Tasmania are intermittently affected by extreme episodes of vegetation fire smoke. The authors are currently pooling hospital admissions and mortality data from 1994 through 2007 for these locations and linking them to historical evidence of large fire events and PM monitoring data to quantify the pollution-related health effects of these events.

This paper presents approaches the authors have developed to identify historical spikes in PM concentrations and to evaluate whether they were caused by vegetation fire smoke or by other possible sources (i.e., dust storms or industrial fires). These methods provide a systematic framework for retrospective identification of the air quality impacts of biomass smoke in a region that is seasonally affected by fires. This approach anticipates increasing impacts on air quality if the frequency and severity of bushfires continues to increase as predicted by the Intergovernmental Panel on Climate Change.[28] This framework could be useful for other epidemiologists working in similar fire-prone environments to contribute to building a useful database for improving the understanding of the public health effects of vegetation fire smoke.

METHODS

Atmospheric PM concentrations measured in six cities from the following three distinct ecoclimatic regions of Australia were studied:

(1) Central New South Wales—a subtropical region with a fire-prone sclerophyll landscape composed of eucalypt forest and smaller patches of headlands (Sydney, Newcastle. and Wollongong). Hot summers, mild winters, and year-round precipitation characterize the climate of this region. These cities experience fire danger in the high, very high, or extreme categories on 7.1% of days.
(2) The Southwest—a region with a semi-arid to Mediterranean climate characterized by long hot summers, mild winters, and limited precipitation during the summer (Perth). Vegetation ranges from closed eucalypt forest to extensive heathland. Perth experiences fire danger in the high, very high, or extreme categories on 33.1% of days.
(3) Tasmania is an island with a temperate climate where vegetation growth is limited in the winter months because of low temperatures, but summers are mild and wet (Hobart and Launceston). This region features areas of tall, closed eucalypt forests carrying high fuel loads. These cities experience fire danger in the high, very high, or extreme categories on 5% of days.

All cities are shown in Figure 1.

For each location, up to 13 yr (between 1994 and 2007) of daily air quality data measured as PM less than 10 μm (PM_{10}) or less than 2.5 μm ($PM_{2.5}$) in aerodynamic diameter were examined. Air pollution data were provided by government agencies in the states of Western Australia, New South Wales, and Tasmania. Daily averages for each site were calculated excluding days with less than 75% of hourly measurements. In Sydney and Perth, where data were collected from several monitoring stations, the missing daily site-specific PM_{10} and $PM_{2.5}$ concentrations were imputed using available data from other proximate monitoring sites in the network. The daily city-wide PM_{10} and $PM_{2.5}$ concentrations were then estimated following the protocol of the Air Pollution and Health: a European Approach studies.[29] For each city, all days in which PM_{10} or $PM_{2.5}$ exceeded the 95th percentile were identified over the entire time series. These extreme values were termed "events." A range of sources was examined to identify the cause of particulate air pollution events, including electronic news archives, Internet searches for other reports, government and research agencies, and satellite imagery. Also examined were remotely sensed aerosol optical thickness (AOT) data to provide further information about days for which the other methods yielded uncertain or conflicting results.

Electronic News Archives

Media searches included the Google News Archive, HighBeam, and Australia-specific reports from News Limited and the Australian Broadcasting Corporation. The following protocol was used:

(1) The search dates were defined as the day before event until 2 days after the event.
(2) Search terms were included to identify each date, location, and potential event types. For example, for events in Sydney, the terms "Sydney," "New South Wales," and "NSW" were included. For event type, the terms "bushfire," "fire," "burning," "smoke," "haze," "air pollution," "particulate," "smog," and "dust" were included.
(3) The URL of relevant articles was documented and the reports were electronically archived.

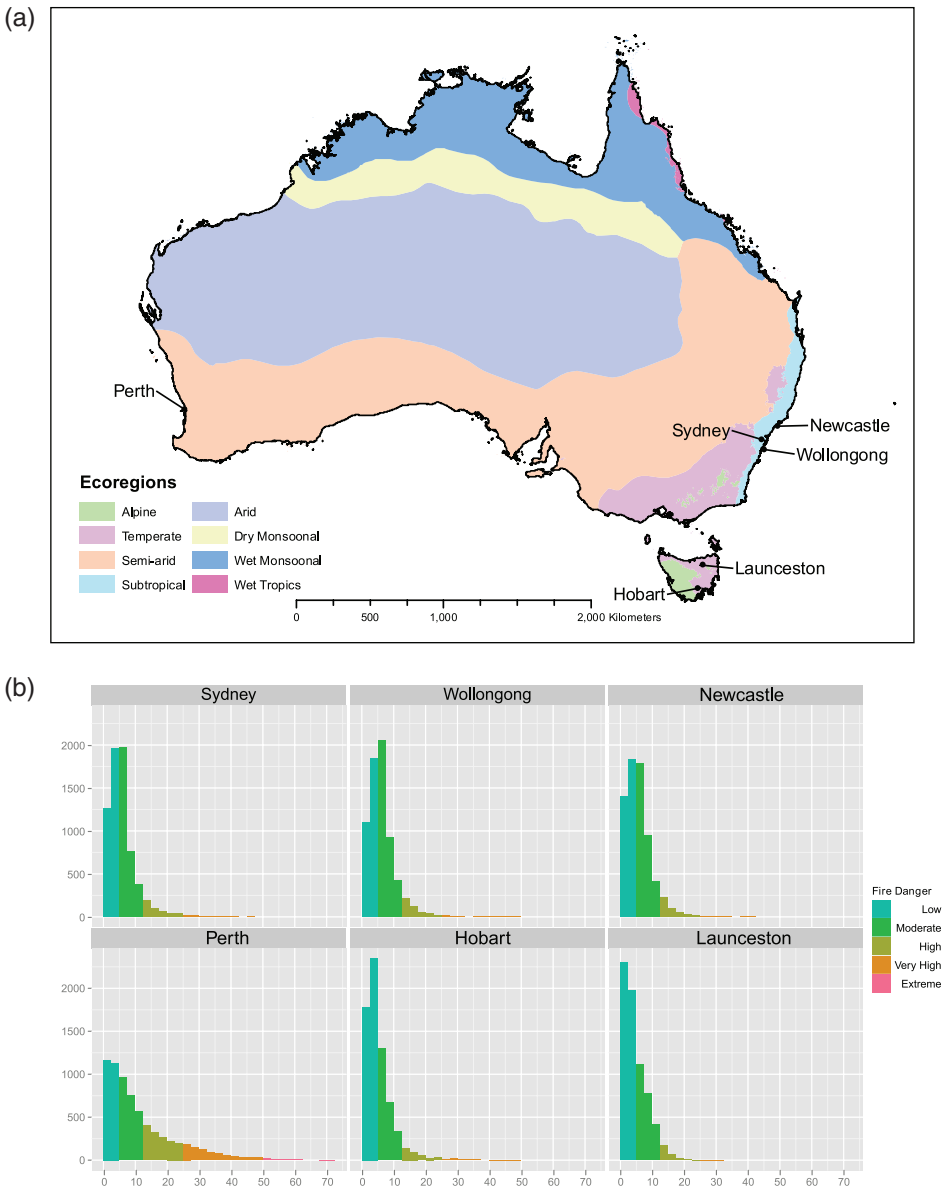


Figure 1. (a) Map of Australia's ecoclimatic regions (based on Hutchinson et al.[41]) showing the location of the six cities in the study. (b) Distribution of McArthur Forest Fire Danger Index MK5 fire danger days calculated from daily meteorological records obtained from the Australian Bureau of Meteorology for each city for the period 1990–2008.

An example of a retrieved news report is reproduced in Figure 2.

Internet Search for Other Reports

The same dates and search terms were also used in a standard Google search to find reports from official bodies such as government departments, weather, or fire services. These were also archived.

Land Management and Research Agencies

The government agencies responsible for land management and bushfire/dust research in each jurisdiction were

*"The Rural Fire Service (RFS) in **New South Wales** says a **bushfire** burning on the central coast is generating a lot of **smoke** but is no cause for alarm. The RFS says it has received many calls from people concerned about smoke on the central coast and in **Sydney**. Crews are backburning a 120 hectare fire at Gwandalan near Wyong. It is expected to be contained overnight and is not threatening property. A minor fire at Albury on the Victorian and **New South Wales** border is also expected to be contained over the next 24 hours."*

www.abc.net.au/news/australia/nsw/newcastle/200501/s1280921.htm

Figure 2. An example of a retrieved news report.

contacted directly and asked to provide information from their internal records. A list of event dates for each PM type was sent to their representatives with a request to identify whether the following types of pollution had been recorded on those dates: (1) bushfire (2) prescribed burn, (3) woodsmoke from domestic wood heater use, and (4) other (including salt and dust). All correspondence was retained in the project archives. Information was provided by the Western Australian Department of Environment and Conservation, the New South Wales Department of Environment, the University of Wollongong Centre for Environmental Risk Management of Bushfires,[30] and the Griffith University Dustwatch Project.[15,16]

Moderate Resolution Imaging Spectroradiometer Satellite Imagery

The Moderate Resolution Imaging Spectroradiometer (MODIS) instruments aboard the National Aeronautics and Space Administration's Aqua and Terra satellites image most of Australia 4 times daily. Data from the red, green, and blue wavelength channels are processed as true color images in near real time. They have been made publicly available since 2001 and are searchable by date via the MODIS Rapid Response System online archives.[31] Thermal anomalies, or "hot spots" (i.e., fires burning on the Earth's surface) are detected by processing data from the 4- and 11-μm wavelengths.[32] Hot spots are marked on the MODIS images; hot spots next to visible smoke plumes represent strong evidence for bushfire-related pollution without the need for great technical expertise. Other causes of hot spots include industrial activities such as power stations; however, these are easy to identify because they remain at a constant location through time on satellite images, whereas bushfires vary in space and time. Although this study covers the period between 1994 and 2007, MODIS true color images were only available from 2001 forward.

The region of Australia was examined on the day of and the day before air pollution episodes. The location of all hot spots and those hot spots with visible smoke plumes affecting the urban center of interest were examined. If smoke plumes were observed to be affecting the city of interest, the URL of the MODIS image was recorded. Satellite images demonstrating hot spots and bushfire smoke affecting Sydney are reproduced in Figure 3.

Total Ozone Mapping Spectrometer

AOT data from the Total Ozone Mapping Spectrometer (TOMS) were examined as a secondary source of information for a subset of days in which media reports were unclear about the location of fires, MODIS imagery was unavailable or inadequate, and other sources of information were lacking. AOT is a unitless measure of light extinction because of aerosol in cloudless atmospheric columns, so it provides evidence of the geographic extent of increased aerosol in the atmosphere. The TOMS AOT data were chosen over AOT data from MODIS because they provide a longer record (1979–1994 and 1994–present) and the data are readily available at the TOMS website.[33] By plotting AOT values in a geographic information system, whether aerosols were concentrated over desert or vegetated regions of the continent was able to be assessed. As such, events could be classified as "dust" or "probable landscape fire smoke," respectively, as described below.

Landscape Fire Smoke. The primary aim was to determine which episodes of particulate air pollution could be confidently ascribed to smoke from landscape fires. Landscape fires from prescribed burning were documented when the information was available, but they were otherwise classified as bushfire. The following were accepted as evidence of landscape fire smoke (LFS): (1) media or government agency reports of fires in or adjacent to cities experiencing peaks in air pollution, and (2) MODIS imagery demonstrating hot spots with associated smoke plumes affecting the urban airshed. Although MODIS imagery provided a high level of confidence for the classification of LFS, absence of evidence from MODIS cannot be equated to an absence of landscape fire. Cloud cover can

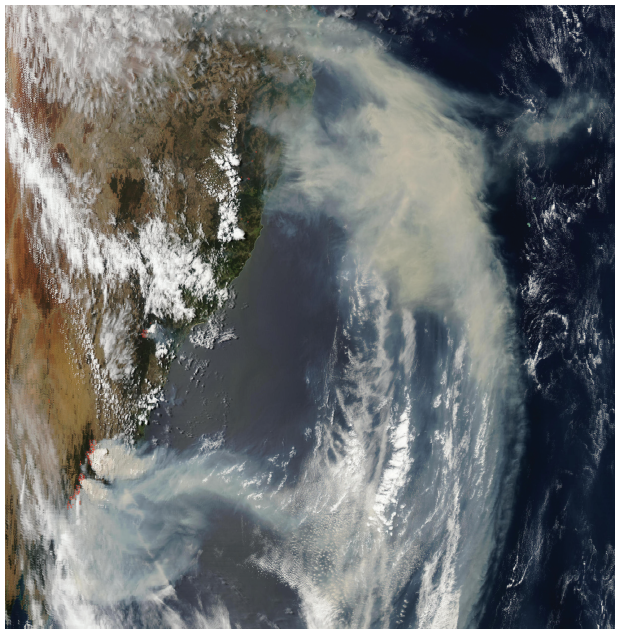


Figure 3. MODIS imagery showing hot spots and smoke plumes affecting Sydney (see bottom left corner of image).

suppress the detection of thermal anomalies and can obscure smoke plumes in true color images. Furthermore, MODIS can miss any fires that occur between the four daily overpasses.

Probable LFS. This category described days for which remote sensing data were highly suggestive of LFS but no confirmatory information was available from other sources. For example, if there were hot spots in adjacent bush land but no smoke plumes were visible on MODIS imagery, the cause of the PM elevation was classified as probable LFS. Similarly, if the TOMS AOT data showed aerosol centered over a vegetated region this was considered likely to be due to landscape fire activity.

Dust. Large dust storms were identified by nearly all of the sources of information. They often generated news stories, were documented by local environment agencies, and were clearly visible in satellite imagery and TOMS AOT data.

Woodsmoke. The authors did not have a systematic method for retrospectively identifying days in which high levels of particulate air pollution could be attributed to the use of wood for domestic heating. However, the internal records of some government agencies did identify some days for which woodsmoke was considered to be the main cause. It is probable that many days of excess pollution in winter in the colder parts of Australia were due to increased use of wood heaters.[34,35]

Nonbiomass Particulate Air Pollution. In some cities, peaks in air pollution were occasionally attributed to higher levels of background pollution associated, for example, with an atmospheric temperature inversion. The media occasionally reported such events.

Nonbiomass Fire Smoke. This category referred to urban factory or warehouse fires and was usually clearly identified as such by media and government agencies.

No Cause Identified. Occasionally no cause could be found for the observed peaks in air pollution. On such days, the peak could have been due to background PM, sea salt, pollution from the increased use of wood heaters in winter, or a combination of factors or specific events such as bushfires or dust storms for which no reports were found.

All primary data sources were independently re-examined on dates for which (1) the available information was inconclusive or conflicting, (2) no event could be identified, and (3) event types were only identified by media sources. A random subset of 5% of all dates was also re-examined. For days in which the information remained inconclusive, the TOMS AOT database was also checked. For example, if there were media reports of fires burning several hundred kilometers from a city, examination of the TOMS database could usefully confirm if the geographic distribution of the aerosols generated by those fires included the city.

The references for the sources of evidence for each date were entered in a Microsoft Access database, including the name of the source (i.e., journal/newspaper/database), the date of publication, and the URL if relevant. Each event was classified as listed below. If there was good evidence for more than one source of air pollution (e.g., bushfire smoke and dust), then both were recorded.

RESULTS

The PM data are summarized in Table 1. The identified causes of these pollution events by city, PM type, and percentile group are shown in Table 2. The distribution of pollution events by season and PM size class is shown in Table 3.

There was considerable variation in the magnitude and seasonal distribution of pollution events between the three ecoregions studied. The three cities in central New South Wales all had the greatest maximum PM events and these occurred most frequently during the summer months. The 99th percentile was very close to Australia's 24-hr air quality standard of 50 $\mu g/m^3$ for PM_{10} in all three cities in this region and to the current reporting standard of 25 $\mu g/m^3$ for $PM_{2.5}$. In the city of Perth in southwestern Australia, peak PM concentrations were not as high: the 99th percentile was 37.4 $\mu g/m^3$ for PM_{10} and 21.4 $\mu g/m^3$ for $PM_{2.5}$. Although PM_{10} events were more frequent in summer, $PM_{2.5}$ events were most frequent in spring. In Tasmania, the

Table 1. Air quality data availability and distribution for the six cities included in this study.

City	PM Type	Data Collected	Number of Days	PM Percentile Distribution ($\mu g/m^3$)							
				Min	25th	50th	75th	95th	97th	99th	Max
Sydney	PM_{10}	1994–2007	5113	1.6	12.8	16.9	21.9	32.0	35.5	47.3	199.2
	$PM_{2.5}$	1996–2007	4256	2.1	6.9	9.1	12.1	18.2	20.3	27.4	100.2
Wollongong	PM_{10}	1994–2007	5068	2.5	12.7	17.2	23.3	35.8	39.8	50.6	280.5
	$PM_{2.5}$	1998–2007	3593	1.6	6.5	8.3	10.9	18.8	18.8	24.6	112.2
Newcastle	PM_{10}	1994–2007	5081	2.0	13.5	17.6	23.2	35.3	38.9	49.7	160.9
	$PM_{2.5}$	1996–2001	4213	1.8	6.8	8.7	11.4	19.2	19.2	25.1	61.9
Perth	PM_{10}	1997–2007	3875	2.6	12.8	16.3	20.5	29.4	31.6	37.4	68.9
	$PM_{2.5}$	1994–2007	5054	0.9	6.1	7.6	9.8	14.8	16.4	21.4	41.5
Hobart	PM_{10}	2006–2007	619	2.1	9.5	13.7	20.1	33.4	35.3	42.4	51.0
	$PM_{2.5}$	2006–2007	575	0.8	3.3	5.1	9.4	22.5	22.5	26.7	31.6
Launceston	PM_{10}	2001–2007	2436	1.0	9.2	14.5	24.3	48.4	54.3	67.7	91.8
	$PM_{2.5}$	2005–2007	941	1.0	4.4	6.5	12.3	33.1	33.1	40.1	55.9

Table 2. A description of all identified events by city, PM size class, and percentile bracket (95–96%, 97–98%, ≥99%).

	Percentile Group							
	PM_{10}				$PM_{2.5}$			
	≥99%	97–98%	95–96%	All ≥95%	≥99%	97–98%	95–96%	All ≥95%
Sydney								
n	52	102	102	256	42	84	84	210
Cause identified	52	63	61	176	39	43	24	106
LFS	46	54	49	149	38	37	21	96
Probable LFS	0	3	9	12	0	4	1	5
Woodsmoke	0	0	0	0	0	0	0	0
Dust	6	2	0	7	1	2	0	3
Nonbiomass fires	1	7	4	12	2	4	2	8
Nonbiomass and not fires	0	3	4	7	2	1	1	4
Newcastle								
n	50	100	100	250	41	81	81	203
Cause identified	34	49	45	128	37	41	27	105
LFS	33	44	34	111	36	39	20	95
Probable LFS	0	5	11	16	1	2	1	10
Woodsmoke	0	0	0	0	0	0	0	0
Dust	1	0	0	3	1	0	1	2
Nonbiomass fires	0	0	0	0	0	0	0	0
Nonbiomass and not fires	0	0	0	0	0	0	0	0
Wollongong								
n	48	96	96	240	36	71	71	178
Cause identified	30	40	27	97	27	29	18	74
LFS	25	27	19	71	25	21	16	62
Probable LFS	4	13	8	25	2	8	2	12
Woodsmoke	0	0	0	0	0	0	0	0
Dust	2	0	1	3	1	1	0	2
Nonbiomass fires	0	0	0	0	0	0	0	0
Nonbiomass and not fires	0	0	0	0	0	0	0	0
Perth								
n	39	76	76	191	51	100	101	252
Cause identified	38	43	35	116	23	17	14	54
LFS	26	35	25	86	17	13	11	41
Probable LFS	2	6	10	18	1	1	2	4
Woodsmoke	10	2	0	12	5	3	1	9
Dust	0	0	0	0	0	0	0	0
Nonbiomass fires	0	0	1	1	0	0	0	0
Nonbiomass and not fires	0	0	0	0	0	0	0	0
Hobart								
n	7	12	12	31	6	10	11	27
Cause identified	3	1	0	4	0	0	1	1
LFS	3	1	NA	4	NA	NA	1	1
Probable LFS	0	0	NA	0	NA	NA	0	0
Woodsmoke	0	0	NA	0	NA	NA	0	0
Dust	0	0	NA	0	NA	NA	0	0
Nonbiomass fires	0	0	NA	0	NA	NA	0	0
Nonbiomass and not fires	0	0	NA	0	NA	NA	0	0
Launceston								
n	24	46	47	117	9	17	17	43
Cause identified	0	0	0	0	4	1	0	5
LFS	NA	NA	NA	NA	4	1	NA	5
Probable LFS	NA	NA	NA	NA	0	0	NA	0
Woodsmoke	NA	NA	NA	NA	0	0	NA	0
Dust	NA	NA	NA	NA	0	0	NA	0
Nonbiomass fires	NA	NA	NA	NA	0	0	NA	0
Nonbiomass and not fires	NA	NA	NA	NA	0	0	NA	0

Notes: NA = not applicable.

99th percentile was 42.4 $\mu g/m^3$ for PM_{10} and 26.7 $\mu g/m^3$ for $PM_{2.5}$ in Hobart, whereas in Launceston the corresponding figures were 67.7 $\mu g/m^3$ for PM_{10} and 40.1 $\mu g/m^3$ for $PM_{2.5}$. Most PM events occurred in winter,[36] which is consistent with the well-documented issue of air pollution from wood heaters in this region.

Table 3. Seasonal distribution of days in which PM_{10} and $PM_{2.5}$ equaled or exceeded the 95th percentile.

	Percent of Days ≥95th Percentile							
	Summer		Autumn		Winter		Spring	
City	PM_{10}	$PM_{2.5}$	PM_{10}	$PM_{2.5}$	PM_{10}	$PM_{2.5}$	PM_{10}	$PM_{2.5}$
Sydney	43	24	14	19	10	27	33	27
Wollongong	48	41	14	18	5	5	33	36
Newcastle	32	33	16	14	11	14	42	38
Perth	52	24	13	33	13	27	22	14
Hobart	14	5	25	25	53	70	8	5
Launceston	1	5	25	14	75	81	1	0

The ability to identify the causes of air pollution events varied by city. It was greatest for the mainland cities of Sydney and Perth and lowest for the Tasmanian cities of Hobart and Launceston (Table 2). There were several reasons for a lower identification rate in Tasmania, including the shorter period of air pollution records and more prevalent use of wood heaters in winter. Tasmania has a colder climate than mainland Australia and, contrary to the other cities examined, most pollution events occurred during winter when bushfires are infrequent. Furthermore, satellite imagery is a less useful resource in Tasmania because of the higher proportion of days for which cloud cover reduces the ability to interpret the satellite images (Table 3).

As expected, the more severe the pollution, the more likely a cause was to be clearly identified. For example, in the city of Sydney there were 52 of 5113 days during which PM_{10} reached or exceeded the 99th percentile. A cause was identified for 48 (92%) of these days, of which 47 (98%) were bushfire smoke and 1 was dust. Of the 256 days on which PM_{10} reached or exceeded the 95th percentile, 172 (67%) had a cause identified and 162 (94%) were attributed to bushfire smoke.

Over the entire study period, the most frequent source of information about air pollution events was from satellite imagery (Table 4). However, before the start of the MODIS archive in 2001, media reports were the primary source of information and the proportion of events for which a cause could be identified was just 30%. This rose to 75% once satellite imagery became available (Table 4.)

DISCUSSION

The main cause of excess particulate pollution episodes in mainland Australian cities was smoke from landscape fires. In Sydney this ranged from 59% of days in which PM_{10} exceeded the 95th percentile to 90% of days in which PM_{10} exceeded the 99th percentile. These were mainly identified from media reports and, from 2001 onward, MODIS satellite imagery. Correlation was high between satellite imagery and media reports for identifying severe bushfire episodes, whereas smaller fires were less frequently reported in the media. The use of multiple sources of information to identify severe pollution episodes gives confidence in the findings, with more than half of the identified events being confirmed by more than one source (Table 5).

The use of satellite imagery had some inherent limitations. These include bad images, cloudy days, and the complete absence of an online MODIS archive before 2001. Although very severe bushfires and pollution episodes were reported in the media, MODIS was the primary source of information and the identification success rate rose considerably after 2001. The TOMS AOT data were a useful adjunct, typically for determining the geographic area affected by smoke from fires identified through media or other reports. The search was restricted to remote sensing data from MODIS and TOMS because both are conveniently available online. However, data from other platforms (e.g., Landsat) or smoke dispersal modeling tools such as the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) might improve future work in this area.[36]

This study was not designed to identify excess pollution from other sources such as motor vehicles or wood heaters, although any such information was recorded if it was reported. Thus, many air pollution episodes from the colder parts of Australia, particularly Tasmania where

Table 4. Proportion of Sydney PM_{10} events due to LFS identified by each source of information before and after the online MODIS archive became available in 2001.

Sources	PM_{10} Events Attributed to LFS 1994–2000 (Pre-MODIS) (n = 30 of 102)	PM_{10} Events Attributed to LFS 2001–2007 (Post-MODIS) (n = 132 of 176)
MODIS	Nil	92
Media	14	56
Government or research agency	7	53
TOMS	2	4

Table 5. Contribution of different sources of information for identifying PM_{10} events in Sydney (2001 onward).

Sources	LFS	Possible LFS	Woodsmoke	Dust	Other Nonbiomass Sources
MODIS only	7	13	0	0	0
Media only	33	0	0	2	17
Govt/Uni only	4	0	0	4	0
TOMS only	0	2	0	0	0
Media + MODIS	22	0	0	0	0
TOMS + Media	1	0	0	0	0
Govt/Uni + MODIS	16	0	0	0	0
Govt/Uni + Media				2	
Govt/Uni + Media+ MODIS	33	0	0	0	0
Media + MODIS + TOMS	1	0	0	0	0
Number of events	117	15	0	6	17

Notes: Govt/Uni = information provided by land management agencies, government departmental reports, or university centers for dust or bushfire research.

winter woodsmoke pollution is a well-recognized problem,[37] were not identified. However, the authors are confident most episodes of LFS and dust storms were found unless these events were so small or geographically localized that they did not cause the citywide PM to exceed the 95th percentile.

The main advantage in this approach was that smaller scale events (i.e., those causing PM elevations between the 95th and 99th percentiles) could be identified that are less likely to attract media attention. Using news archives and government records, Morgan et al.[2] were able to attribute 32 pollution episodes in Sydney to bushfire smoke. However, analyses were restricted to events in which PM_{10} was greater than the 99th percentile and thus likely to be reported by news media. Over the same period an additional 3 days over the 99th percentile and a further 120 days in which LFS caused the PM_{10} to reach the 95th to 99th percentile were identified.

LFS is an episodic and increasingly important environmental hazard for human populations. Droughts, extreme fire weather, and landscape fire activity are increasing globally, and fire emissions are an important source of greenhouse gases.[38] This database of identified episodes has been developed as an ongoing resource for future environmental epidemiologic studies, including LFS episodes and other related exposures such as heat waves and dust storms. The database is a step forward in overcoming one of the major limitations of studies of the health effects of exposure to LFS to date—accurate exposure assessment. These results have now been used to undertake the first Australian study of mortality associated with pollution events from bushfires and dust storms.[39] The database will be refined and improved over time, and information from other sources (e.g., the Australasian Fire Authorities Council's Australian Incident Reporting System[40]) can be added.

The manual searching of multiple information sources adds value to these findings. However, it is labor-intensive and difficult to automate. As an extension of this project, the authors have been using MODIS measurements on remotely sensed fires to model their relationship with air quality measurements and evaluate their utility for identifying pollution episodes from landscape fires. The database described here has allowed the authors to validate this less-intensive approach and to apply it within different settings in Australia and Canada. The authors hope that others working in similar regions will consider using similar methods to assess their historical data so that, together, an extensive validated PM event database can be developed for health and landscape research that will continue to improve global understanding of the public health effects of vegetation fire smoke.

REFERENCES

1. Hänninen, O. Challenges in Estimating the Health Effects of Biomass Smoke. Response to Sverre Vedal and Steven J. Dutton: Wildfire Air Pollution and Daily Mortality in a Large Urban Area; *Environ. Res.* **2008**, *106*, 423-425; discussion 425.
2. Morgan, G.; Sheppeard, V.; Khalaj, B.; Ayyar, A.; Lincoln, D.; Jalaludin, B.; Beard, J.; Corbett, S.; Lumley, T. The Effects of Bushfire Smoke on Daily Mortality and Hospital Admissions in Sydney, Australia, 1994 to 2002; *Epidemiology* **2010**, *21*, 47-55.
3. Duclos, P.; Sanderson, L.M.; Lipsett, M. The 1987 Forest Fire Disaster in California: Assessment of Emergency Room Visits; *Arch. Environ. Health* **1990**, *45*, 53-58.
4. Chen, L.; Verrall, K.; Tong, S. Air Particulate Pollution Due to Bushfires and Respiratory Hospital Admissions in Brisbane, Australia; *Int. J. Environ. Health Res.* **2006**, *16*, 181-191.
5. Arbex, M.A.; Martins, L.C.; de Oliveira, R.C.; Pereira, L.A.A.; Arbex, F.F.; Cancado, J.E.D.; Saldiva, P.H.N.; Braga, A.L.F. Air Pollution from Biomass Burning and Asthma Hospital Admissions in a Sugar Cane Plantation Area in Brazil; *J. Epidemiol. Community Health* **2007**, *61*, 395-400.
6. Delfino, R.J.; Brummel, S.; Wu, J.; Stern, H.; Ostro, B.; Lipsett, M.; Winer, A.; Street, D.H.; Zhang, L.; Tjoa, T. The Relationship of Respiratory and Cardiovascular Hospital Admissions to the Southern California Wildfires of 2003; *Occup. Environ. Med.* **2009**, *66*, 189-197.
7. Frankenberg, E.; McKee, D.; Thomas, D. Health Consequences of Forest Fires in Indonesia; *Demography* **2005**, *42*, 109-129.
8. Golshan, M.; Faghihi, M.; Roushan-Zamir, T.; Marandi, M.M.; Esteki, B.; Dadvand, P.; Farahmand-Far, H.; Rahmati, S.; Islami, F. Early Effects of Burning Rice Farm Residues on Respiratory Symptoms of Villagers in Suburbs of Isfahan, Iran; *Int. J. Environ. Health Res.* **2002**, *12*, 125-131.
9. Hanigan, I.C.; Johnston, F.H.; Morgan, G.G. Vegetation Fire Smoke, Indigenous Status and Cardio-Respiratory Hospital Admissions in Darwin, Australia, 1996–2005; *BMC Environ. Health* **2008**, *5*, 42.
10. Johnston, F.H.; Bailie, R.; Pilotto, L.; Hanigan, I. Ambient Biomass Smoke and Cardio-Respiratory Hospital Admissions; *BMC Public Health* **2007**, *7*, 240.
11. Johnston, F.H.; Webby, R.J.; Pilotto, L.; Bailie, R.; Parry, D.; Halpin, S. Vegetation Fires, Particulate Air Pollution and Asthma: a Panel Study in the Australian Monsoon Tropics; *Int. J. Environ. Health Res.* **2006**, *16*, 391-404.
12. Kunii, O.; Kanagawa, S.; Yajima, I.; Hisamatsu, Y.; Yamamura, S.; Amagai, T.; Ismail, I.T. The 1997 Haze Disaster in Indonesia: Its Air Quality and Health Effects; *Arch. Environ. Health* **2002**, *57*, 16-22.

13. Kunzli, N.; Avol, E.; Wu, J.; Gauderman, W.J.; Rappaport, E.; Millstein, J.; Bennion, J.; McConnell, R.; Gilliland, F.D.; Berhane, K.; Lurmann, F.; Winer, A.; Peters, J.M. Health Effects of the 2003 Southern California Wildfires on Children; *Am. J. Respir. Crit. Care Med.* **2006**, *174*, 1221-1228.
14. Moore, D.; Copes, R.; Fisk, R.; Joy, R.; Chan, K.; Brauer, M. Population Health Effects of Air Quality Changes Due to Forest Fires in British Columbia in 2003: Estimates from Physician-Visit Billing Data; *Can. J. Public Health* **2006**, *97*, 105-108.
15. Mott, J.A.; Mannino, D.M.; Alverson, C.J.; Kiyu, A.; Hashim, J.; Lee, T.; Falter, K.; Redd, S.C. Cardiorespiratory Hospitalizations Associated with Smoke Exposure during the 1997 Southeast Asian Forest Fires; *Int. J. Hygiene Environ. Health* **2005**, *208*, 75-85.
16. Mott, J.A.; Meyer, P.; Mannino, D.; Redd, S.C.; Smith, E.M.; Gotway-Crawford, C.; Chase, E. Wildland Forest Fire Smoke: Health Effects and Intervention Evaluation, Hoopa, California, 1999; *West. J. Med.* **2002**, *176*, 162-163.
17. Ovadnevaitd, J.; Kvietkus, K.; Marsalka, A. 2002 Summer Fires in Lithuania: Impact on the Vilnius City Air Quality and the Inhabitants Health; *Sci. Total Environ.* **2006**, *356*, 11-21.
18. Sutherland, E.R.; Make, B.J.; Vedal, S.; Zhang, L.; Dutton, S.J.; Murphy, J. R.; Silkoff, P.E. Wildfire Smoke and Respiratory Symptoms in Patients with Chronic Obstructive Pulmonary Disease; *J. Allergy Clin. Immunol.* **2005**, *115*, 420-422.
19. Tham, R.; Erbas, B.; Akram, M.; Dennekamp, M.; Abramson, M.J. The Impact of Smoke on Respiratory Hospital Outcomes during the 2002–2003 Bushfire Season, Victoria, Australia; *Respirology* **2009**, *14*, 69-75.
20. Naeher, L.P.; Brauer, M.; Lipsett, M.; Zelikoff, J.T.; Simpson, C.D.; Koenig, J. Q.; Smith, K.R. Woodsmoke Health Effects: a Review; *Inhal. Toxicol.* **2007**, *19*, 67-106.
21. Pope, C.A.; Dockery, D.W. Critical Review. Health Effects of Fine Particulate Air Pollution: Lines That Connect; *J. Air & Waste Manage. Assoc.* **2006**, *56*, 709-742.
22. Sastry, N. Forest Fires, Air Pollution, and Mortality in Southeast Asia; *Demography* **2002**, *39*, 1-23.
23. Vedal, S.; Dutton, S.J. Wildfire Air Pollution and Daily Mortality in a Large Urban Area; *Environ. Res.* **2006**, *102*, 29-35.
24. Hanninen, O.O.; Salonen, R.O.; Koistinen, K.; Lanki, T.; Barregard, L.; Jantunen, M. Population Exposure to Fine Particles and Estimated Excess Mortality in Finland from an East European Wildfire Episode; *J. Expo. Sci. Environ. Epidemiol.* **2009**, *19*, 414-422.
25. Jalava, P.I.; Salonen, R.O.; Hälinen, A.I.; Penttinen, P.; Pennanen, A.S.; Sillanpää, M.; Sandell, E.; Hillamo, R.; Hirvonen, M.R. In Vitro Inflammatory and Cytotoxic Effects of Size-Segregated Particulate Samples Collected during Long-Range Transport of Wildfire Smoke to Helsinki; *Toxicol. Appl. Pharmacol.* **2006**, *215*, 341-353.
26. Danielsen, P.H.; Loft, S.; Kocbach, A.; Schwarze, P.E.; Møller, P. Oxidative Damage to DNA and Repair Induced by Norwegian Wood Smoke Particles in Human A549 and THP-1 Cell Lines; *Mutat. Res.* **2009**, *674*, 116-122.
27. Mazzoli-Rocha, F.; Magalhães, B. Comparative Respiratory Toxicity of Particles Produced by Traffic and Sugar Cane Burning; *Environ. Res.* **2008**, *108*, 35-41.
28. Confalonieri, U.; Menne, B.; Akhtar, R.; Ebi, K.L.; Hauengue, M.; Kovats, R.S.; Revich, B.; Woodward, A. In *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*; Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E., Eds.; Cambridge University: Cambridge, United Kingdom, 2007; p 402.
29. Atkinson, R.W.; Anderson, H.R.; Sunyer, J.; Ayres, J.; Baccini, M.; Vonk, J.M.; Boumghar, A.; Forastiere, F.; Forsberg, B.; Touloumi, G.; Schwartz, J.; Katsouyanni, K. Acute Effects of Particulate Pollution on Respiratory Admissions. Results from the APHEA 2 Project; *Am. J. Respir. Crit. Care Med.* **2001**, *164*, 1860-1866.
30. *Centre for Environmental Risk Management of Bushfires*; University of Wollongong; available at *http://www.uow.edu.au/science/biol/icb/cermb.html* (accessed 2010).
31. *MODIS Rapid Response System*; National Aeronautics and Space Administration; available at *http://rapidfire.sci.gsfc.nasa.gov* (accessed October 2010).
32. Kaufman, Y.J.; Ichoku, C.; Giglio, L.; Korontzi, S.; Chu, D.A.; Hao, W.M.; Li, R.R.; Justice, C.O. Fire and Smoke Observed from the Earth Observing System MODIS Instrument: Products, Validation, and Operational Use; *Int. J. Remote Sens.* **2003**, *24*, 1765-1781.
33. *Total Ozone Mapping Spectrometer*; National Aeronautics and Space Administration; available at *http://toms.gsfc.nasa.gov/aerosols/aerosols_v8.html* (accessed October 2010).
34. Hinwood, A.L.; De Klerk, N.; Rodriguez, C.; Jacoby, P.; Runnion, T.; Rye, P.; Landau, L.; Murray, F.; Feldwick, M.; Spickett, J. The Relationship between Changes in Daily Air Pollution and Hospitalizations in Perth, Australia 1992–1998: a Case-Crossover Study; *Int. J. Environ. Health Res.* **2006**, *16*, 27-46.
35. Robinson, D.L.; Monro, J.M.; Campbell, E.A. Spatial Variability and Population Exposure to $PM_{2.5}$ Pollution from Woodsmoke in a New South Wales Country Town; *Atmos. Environ.* **2007**, *41*, 5464-5478.
36. Draxler, R.R.; Hess, G.D. An Overview of the HYSPLIT_4 Modeling System of Trajectories, Dispersion, and Deposition; *Australian Meteorol. Mag.* **1998**, *47*, 295-308.
37. Keywood, M.D.; Ayers, G.P.; Gras, J.L.; Gillett, R.W.; Cohen, D.D. Size Distribution and Sources of Aerosol in Launceston, Australia, during Winter 1997; *J. Air & Waste Manage. Assoc.* **2000**, *50*, 418-427.
38. Bowman, D.M.J.S.; Balch, J.K.; Artaxo, P.; Bond, W.J.; Carlson, J.M.; Cochrane, M.A.; D'Antonio, C.M.; DeFries, R.S.; Doyle, J.C.; Harrison, S.P. Johnston, F.H.; Keeley, J.E.; Krawchuk, M.A.; Kull, C.A.; Marston, J.B.; Moritz, M.A.; Prentice, I.C.; Roos, C.I.; Scott, A.C.; Swetnam, T.W.; van der Werf, G.R.; Pyne, S.J. Fire in the Earth System; *Science* **2009**, *324*, 481-484.
39. Johnston, F.H.; Hanigan, I.C.; Henderson, S.B.; Morgan, G.G.; Bowman, D. M.J.S. The Association between Mortality and Extreme Particulate Air Pollution Events from Bushfires and Dust Storms in Sydney, Australia 1994–2007, submitted for publication, *Environ. Res.*
40. Bryant, C. *Understanding Bushfire: Trends in Deliberate Vegetation Fires in Australia*; Australian Institute of Criminology: Canberra, Australia, 2008; p 56.
41. Hutchinson, M.F.; McIntyre, S.; Hobbs, R.J.; Stein, J.L.; Garnett, S.; Kinloch, J. Integrating a Global Agro-Climatic Classification with Bioregional Boundaries in Australia; *Global Ecol. Biogeography* **2005**, *14*, 197-212.

About the Authors

Fay H. Johnston is a public health physician and research fellow and Sarah B. Henderson is a research fellow at the Menzies Research Institute at the University of Tasmania. Ivan C. Hanigan is a research associate and data manager at the National Centre of Epidemiology and Population Health at the Australian National University in Canberra, Australia. Geoffrey G. Morgan is an associate professor at the North Coast Area Health Service and Department of Rural Health at the University of Sydney in Lismore, New South Wales, Australia. Grant J. Williamson is a research fellow, Talia Portner is a research associate, and David M.J.S. Bowman is a research professor at the School of Plant Sciences at the University of Tasmania. Please address correspondence to: Fay Johnston, Menzies Research Institute, University of Tasmania, Private Bag 23 Hobart 7001, Tasmania, Australia; phone: +61-3-62267726; fax: +61-3-62267755; e-mail: fay.johnston@utas.edu.au.

4.2 Dataset 3: The validated bushfire smoke events database

Title: Hanigan, I. C. (2015). The Validated Bushfire Smoke Events Database. https://gislibrary-extreme-weather.anu.edu.au/biomass_smoke_events

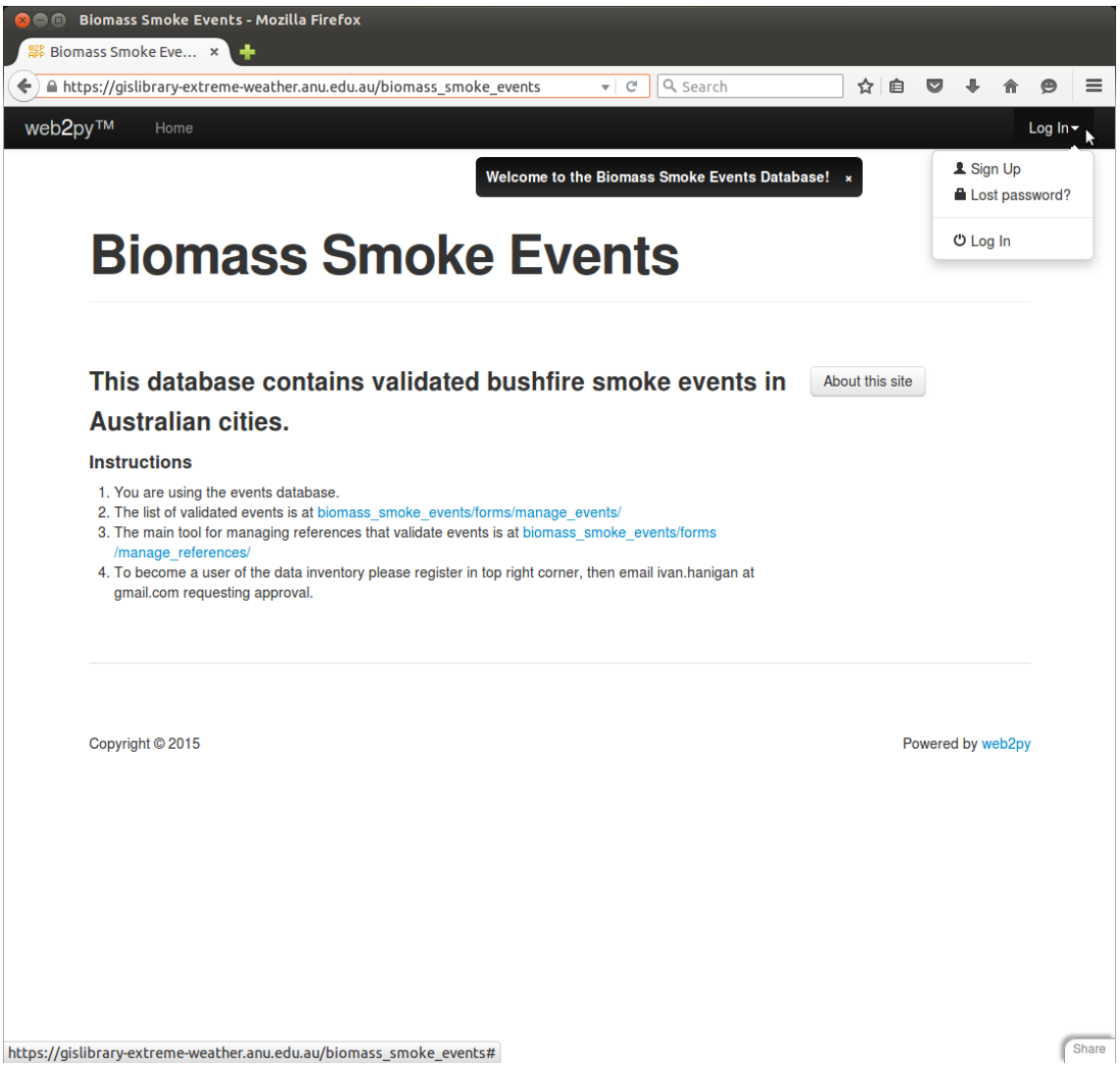
Stage of publication: Published data package at the ANU Data Commons Portal with website front-end.

Journal Impact Factor: NA

My contribution: I was the sole author of this online database server and website front-end. I was responsible for designing and implementing the web-server for the distribution of this database (which was created as per description in papers 4 and 5). I liaised with systems administration staff at the Australian National University Information Technology division to ensure the security of the database server.

Approvals required for inclusion of Dataset 3 in this thesis:

Name	Notes	Signature/date
Mr. Ivan Hanigan		2015-09-28
Copyright	Creative Commons - Attribution 4.0. The licence allows others copy, distribute, display, and perform the work and derivative works based upon it provided that they credit the original source.	
Ethics Approval	Not Required	


Biomass Smoke Events - Mozilla Firefox
Biomass Smoke Eve...
https://gislibrary-extreme-weather.anu.edu.au/biomass_smoke_events
Search
web2py™
Home
Log In
Welcome to the Biomass Smoke Events Database!
Sign Up
Lost password?
Log In
Biomass Smoke Events
This database contains validated bushfire smoke events in Australian cities.
About this site
Instructions
1. You are using the events database.
2. The list of validated events is at biomass_smoke_events/forms/manage_events/
3. The main tool for managing references that validate events is at biomass_smoke_events/forms/manage_references/
4. To become a user of the data inventory please register in top right corner, then email ivan.hanigan at gmail.com requesting approval.
Copyright © 2015
Powered by web2py
https://gislibrary-extreme-weather.anu.edu.au/biomass_smoke_events#
Share

4.3 Software package 2: Code for smoke events database (dataset 3)

Title: Hanigan, I. C. (2015). Biosmoke Validated Events. https://github.com/swish-climate-impact-assessment/BiosmokeValidatedEvents.

Stage of publication: Open-source software available on Github

Journal Impact Factor: NA

My contribution: I was the lead author on this software package. I created the code and tested the functions on multiple datasets representing air pollution across many regions of Australia. I led the creation of the package help manual. I distribute this software on my website.

Approvals required for inclusion of Software package 2 in this thesis:

Name	Notes	Signature/date
Mr. Ivan Hanigan		2015-09-28
Copyright	Creative Commons - Attribution 4.0. The licence allows others copy, distribute, display, and perform the work and derivative works based upon it provided that they credit the original source.	
Ethics Approval	Not applicable	

Package 'BiosmokeValidatedEvents'

November 15, 2015

Title an R Package to reproduce the extreme biomass smoke event database

Version 0.1

Date 2015-07-08

Maintainer ivanhanigan <ivan.hanigan@gmail.com>

Description
This package contains a source for the Australian Biomass Smoke Validated Events Database.

Suggests rpostgrestools,
knitr,
rmarkdown

License CC BY 4.0

LazyData true

VignetteBuilder knitr

R topics documented:

`all_stations_all_dates` *All Stations, All Dates*

Description

All Stations, All Dates

Usage

```
all_stations_all_dates(town, pollutant)
```

Arguments

`town`	Biomass Study area
`pollutant`	you got it

Value

text for a query

`BiosmokeValidatedEvents` *BiosmokeValidatedEvents*

Description

BiosmokeValidatedEvents

`citywide_av` *city wide average*

Description

city wide average

Usage

```
citywide_av(town, poll, stat)
```

Arguments

`town`

`poll`

`stat`

Value

nothing to R, this creates things in the database

`impute` *impute for each site*

Description

impute for each site

Usage

```
impute(sitelist = c("SouthLake", "Duncraig"), town = "PERTH",
  poll = "pm10", stat = "av", maxdate = "2007-12-31")
```

Arguments

`sitelist`	sites
`town`	town
`poll`	pollutant
`stat`	statistical unit as per avg or max
`maxdate`	the end of the time series

Value

database table

`mindates_pollutant` *Minimum Date a Pollutant is observed from*

Description

Minimum Date a Pollutant is observed from

Usage

```
mindates_pollutant(town = "perth", pollutant = "pm10_av")
```

Arguments

`town`	the Biomass Study Town in question
`pollutant`	you got it

Value

text for a SQL query

`missing99` *99th centile missing references of any type*

Description

99th centile missing references of any type

Usage

```
missing99(poll)
```

Arguments

`poll`	pollutant

Value

list of dates

`n_missing` *number missing*

Description

number missing

Usage

```
n_missing(town, poll, thresh = 0.05)
```

Arguments

`town`	the one to do
`poll`	pollutant
`thresh`	theshold below which we will do it

Value

nmissing is a message like 'go for it'

`sites_todo` *sites with potential*

Description

sites with potential

Usage

```
sites_todo(town, mindate, maxdate = "2007-12-31", threshold = 0.7, poll,
  stat)
```

Arguments

`town`

`mindate`

`maxdate`

`threshold`

`poll`

`stat`

Value

text for a sql query

`stitch_together` *put all the bits together*

Description

put all the bits together

Usage

```
stitch_together(poll = polls[5, 3], stat = "av")
```

Arguments

`poll`	pollutant
`stat`	av or max

Value

tables in the database

Index

4.4 Paper 5: The effect on mortality of extreme air pollution events from bushfires and duststorms

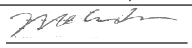
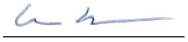
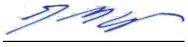
Title: Johnston, F.J., Hanigan, I.C., Henderson, S.B., Morgan, G.G, and Bowman, D.M.J.S. (2011). Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994-2007. Environmental Research 111 (6): 811-816. doi:10.1016/j.envres.2011.05.007

Stage of publication: Published peer-reviewed paper

Journal Impact Factor: 4.4

My contribution: In this paper I worked with my co-authors to design the analysis based on the hypothesis proposed by Dr Johnston and Prof. Morgan. I took responsibility for the data processing stage to create the exposure data, cleaned the health outcome data, estimated the influenza epidemic outbreaks, and cleaned the public holiday data. I conducted all data analyses. I produced figures and summarised the results. I helped to interpret and write about the results and the discussion. I also assisted prepare the text for the methods and results section of the paper. I was responsible author for acquiring ethics committee approval for access to the mortality data.

Approvals required for inclusion of Paper 5 in this thesis:

Name	Notes	Signature/date
Dr. Fay H. Johnston		 2015-08-21
Mr. Ivan Hanigan		 2015-09-28
Dr. Sarah B. Henderson		 2015-10-06
Prof. Geoffrey G. Morgan		 2015-11-15
Prof. David MJS Bowman		 2015-10-06
Copyright	License number: 3723591078860 License date: Oct 07, 2015 Licensed content publisher: Elsevier	2015-10-07
Ethics Approval	ANU Human Research Ethics Protocol: 2008/199. Date entered 26/04/2008, variation = 03/06/2011 (extension of study period, addition of more data).	

Environmental Research 111 (2011) 811–816

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994–2007 ☆

Fay Johnston [a,*], Ivan Hanigan [b], Sarah Henderson [c], Geoffrey Morgan [d], David Bowman [e]

[a] Menzies Research Institute, University of Tasmania, Private Bag 23, Hobart 7001, Tasmania, Australia
[b] School of Plant Science, University of Tasmania and the Australian National University, Corner of Eggleston and Mills Road, Canberra 0200, Australian Capital Territory, Australia
[c] University of Tasmania and the British Colombia Centre for Disease Control, Vancouver, Canada
[d] North Coast Area Health Service and Department of Rural Health, University of Sydney, 55-61 Uralba Street, Lismore, 2480 New South Wales, Australia
[e] School of Plant Sciences, University of Tasmania, Private Bag 50, Hobart 7001, Tasmania, Australia

ARTICLE INFO

Article history:
Received 7 December 2010
Received in revised form
21 April 2011
Accepted 6 May 2011
Available online 23 May 2011

Keywords:
Air pollution
Particulate matter
Bushfire smoke
Dust
Mortality

ABSTRACT

Introduction: Extreme air pollution events due to bushfire smoke and dust storms are expected to increase as a consequence of climate change, yet little has been published about their population health impacts. We examined the association between air pollution events and mortality in Sydney from 1997 to 2004.

Methods: Events were defined as days for which the 24 h city-wide concentration of PM_{10} exceeded the 99th percentile. All events were researched and categorised as being caused by either smoke or dust. We used a time-stratified case-crossover design with conditional logistic regression modelling adjusted for influenza epidemics, same day and lagged temperature and humidity. Reported odds ratios (OR) and 95% confidence intervals are for mortality on event days compared with non-event days. The contribution of elevated average temperatures to mortality during smoke events was explored.

Results: There were 52 event days, 48 attributable to bushfire smoke, six to dust and two affected by both. Smoke events were associated with a 5% increase in non-accidental mortality at a lag of 1 day OR (95% confidence interval (CI)) 1.05 (95%CI: 1.00–1.10). When same day temperature was removed from the model, additional same day associations were observed with non-accidental mortality OR 1.05 (95%CI: 1.00–1.09), and with cardiovascular mortality OR (95%CI) 1.10 (95%CI: 1.00–1.20). Dust events were associated with a 15% increase in non-accidental mortality at a lag of 3 days, OR (95%CI) 1.16 (95%CI: 1.03–1.30).

Conclusions: The magnitude and temporal patterns of association with mortality were different for smoke and dust events. Public health advisories during bushfire smoke pollution episodes should include advice about hot weather in addition to air pollution.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Australia is an arid continent with an ancient history of dust storms and bushfire activity that predates human occupation (Bowman, 2000). Both events cause episodes of severe air pollution and both are set to increase in frequency and severity as global temperatures continue to warm (Confalonieri et al., 2007). The year of 2009 included Australia's worst bushfires and one of the most severe dust events on record, highlighting our particular vulnerability to these manifestations of climate change (Confalonieri et al., 2007). Severe air pollution events from bushfires and dust are usually short-lived, but the resulting particulate matter (PM) can reach extreme concentrations, travel vast distances, and affect densely populated areas far from their source.

Urban air pollution has been causally associated with non-accidental, respiratory and cardiovascular mortality (Pope and Dockery, 2006), yet data concerning association between mortality

☆ *Funding and Ethical Approvals*: The study was supported by an Australian Research Council linkage grant (LP0882048) with contributions from the following industry partners: The Tasmanian Department of Health and Human Services, Environment Protection Authority and Fire Service; the Western Australia Department of Health and Department of Environment and Conservation and the New South Wales Department of Health. Fay Johnston is supported by a fellowship from Australia's National Health and Medical Research Council (No. 490057). The funding agencies had no role in the analysis of data or preparing the manuscript. This research was approved by the human research ethics committees of the University of Tasmania and the Australian National University.

* Corresponding author. Fax: +613 62267755.
E-mail addresses: fay.johnston@utas.edu.au (F. Johnston), ivan.hanigan@anu.edu.au (I. Hanigan), sarah.henderson@bccdc.ca (S. Henderson), geoff.morgan@ncahs.health.nsw.gov.au (G. Morgan), david.bowman@utas.edu.au (D. Bowman).

0013-9351/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.envres.2011.05.007

and severe dust or wildfire smoke events are sparse. A handful of toxicological studies suggests that smoke particles elicit a spectrum of effects similar to those of urban air pollution, including inflammation, oxidative stress and pro-coagulation (Jalava et al., 2006), but epidemiological studies are limited and have reported mixed outcomes. Of five previous studies of mortality and PM from wildfire smoke, only the two conducted during the extreme South East Asian wildfires of 1997–1998 reported a positive association (Jayachandran, 2005; Sastry, 2002). Two further studies from Finland and Sydney both reported increases in mortality similar in magnitude to studies of urban PM, but neither result was statistically significant (Hanninen et al., 2009; Morgan et al., 2010). A small study from Denver, USA found no association with mortality after a 2-day smoke pollution episode, but the statistical power of this study was limited (Hänninen, 2008; Vedal and Dutton, 2006). Studies of mortality in association with extreme dust events are also sparse. Perez et al. (2008) found an association between windblown dust from the Sahara desert and daily mortality in Barcelona, Spain. However, three studies of Asian dust storms had mixed results, with two finding small increases in non-accidental mortality that were not statistically significant (Chen et al., 2004; Hashizume et al., 2010; Kwon et al., 2002).

In summary, there is still insufficient evidence to determine whether the particulate matter from vegetation fire smoke or dust storm pollution carries a greater or lesser risk of mortality compared with PM from urban sources. The episodic nature of large events makes them difficult to study. It is challenging to predict precisely when and where air quality will be affected, so studies are generally retrospective and dependent upon routinely collected data. Further, the peak exposures may be too short-lived to detect all but the most sensitive health outcomes (Hänninen, 2008). Australia provides the opportunity to explore outcomes associated with both these exposures. Australian cities are frequently affected by bushfire smoke given their proximity to highly flammable native vegetation and hot dry weather conditions that favour combustion. The large arid interior of the continent is a major source of global atmospheric dust. Drought, overgrazing and bushfires all contribute to dust storms that occasionally affect Australian cities (Mitchell et al., 2010). Here we retrospectively assess the mortality associated with extreme air pollution events due to bushfire smoke and dust in Sydney (population 3,862,000) from January 1994 to June 2007.

2. Methods

2.1. Exposure definitions

Air quality data were provided by the New South Wales Department of Environment, Climate Change and Water. We obtained thirteen and a half years of daily average concentrations of particulate matter less than 10 μm in aeronautic diameter (PM_{10}). PM_{10} was measured at 7 monitoring stations across the city for which data were available for 70% or greater of days in the time series. All stations used tapered element oscillating microbalances for measuring PM concentrations. Missing data for each site were imputed using available data from other proximate monitoring sites in the network following the protocol of the Air Pollution and Health a European Approach studies in which the missing value is replaced with the weighted average of the values of the rest of the monitoring stations (Atkinson et al., 2001). We then calculated the daily city-wide average PM_{10} concentrations for the study period. Daily measures of 1 h maximum ozone (parts per billion) were provided for 9 stations. City-wide averages were calculated using the same methods described for PM above.

Smoke and dust events were identified from a validated database that is the subject of another paper (Johnston et al., 2011). In brief, an extreme pollution event was defined as any day on which the PM_{10} concentration exceeded the 99th percentile of the time series (47.3 $\mu g/m^3$), and multiple sources of information were used to evaluate the cause of each event. We began by searching news archives for stories related to dust and smoke pollution. Next we searched for evidence of smoke and dust plumes using readily-available remote sensing data from the moderate resolution imaging spectroradiometer and the total ozone mapping spectrometer provided by the US National Aeronautical and Space Administration (NASA, 2010). Finally, Government land management agencies provided us with information about dates of prescribed burning and wildfires, and the Dustwatch project of Griffith University, Brisbane provided information from their Dust Event Database. The latter uses computer simulations to estimate the origin and geographic distribution of Australia dust events (McTainsh, 1998; Raupach et al., 1994).

We selected the 99th percentile as the cut point for defining an extreme air pollution event because it corresponded closely with Australia's 24-h air quality standard for PM_{10} (50 $\mu g/m^3$) and the majority of air quality exceedences in Australia are due to these extreme events. This allows our results to be useful for the public health authorities that are required to act when air quality standards are breached. Previous bushfire smoke research in Sydney has also used the 99th percentile to define severe pollution events (Morgan et al., 2010). As a sensitivity analysis we also conducted analyses including less severe smoke and dust pollution events in which PM_{10} exceeded the 95th percentile (32 $\mu g/m^3$) (Fig. 2).

The 99th percentile was exceeded on 56 occasions in Sydney. All events were identified as being caused by either smoke or dust, with 52 validated by two or more separate sources. When the 95th percentile, was used for the cut point to define an extreme event, only 106 of the 204 exceedences could be confidently attributed to smoke or dust with 43 validated by two or more separate sources (Fig. 2) (Johnston et al., 2011). Days not identified as having bushfire smoke or dust events were classified as background. Sources of background particulate matter in Sydney include motor vehicle and industrial emissions, domestic wood smoke and crustal particles (NSW Department of Environment Climate Change and Water, 2010). In Sydney the vast majority of bushfire smoke is derived from wildfires in the eucalypt forests to the west of the city, while dust is transported from the arid interior of the continent (Fig. 1) (Johnston et al., 2011).

Daily average ambient temperature and humidity (as measured by dew point temperature) were provided by the Bureau of Meteorology and merged with the health data using a technique that weights observations from different weather stations according to population density (Hanigan et al., 2006). Epidemics of influenza were defined as days with hospital admission rates for influenza greater than the 90th percentile of the regional distribution (Touloumi et al., 2005).

2.2. Outcome data

Our study population was defined geographically according to their statistical local area of residence. We included all statistical local areas in the Sydney metropolitan area for which representative air quality monitoring data were available. Mortality data were provided by the Australian Bureau of Statistics coded according to the World Health Organisation's International Classification of Diseases (ICD) (Australian Bureau of Statistics; WHO, 2005). Until December 1996 the underlying cause of death was coded according to the ICD 9th revision (ICD-9) and thereafter according to the 10th revision (ICD-10). We followed the protocol of the National Casemix and Classification Centre for mapping from ICD-9 to ICD-10 (National Casemix and Classification Centre, 2010). We examined non-accidental mortality, (ICD-9 codes <800; ICD-10 codes A00-R99) cardiovascular mortality (ICD-9 codes 390-459; ICD-10 codes I00-I99 excluding 67.3, I68.0, I88, I97.8-98.0, G45 excluding G45.3, G46, M30, M31 and R58) and respiratory mortality (ICD-9 codes 460–519; ICD-10 codes J00-J99 excluding J95.4 to J95.9, R09.1, R09.8).

2.3. Statistical analyses

All analyses were conducted using the R statistical software package (R Foundation for Statistical Computing, 2006). We used a time-stratified case-crossover design in which the event status, coded as a binary indicator variable for smoke and dust, on (and up to three days before) the day of death was compared with the event status on control days matched by day of week, month and calendar year. By design this approach controls for the effects of day of week, season and long term trends on mortality (Maclure, 1991). We used conditional logistic regression models adjusted for meteorology and influenza epidemics to estimate the odds ratio (OR) for mortality associated with smoke and dust events compared with the background. To determine the optimal number of degrees of freedom for modelling the meteorological variables in the logistic models, we first conducted Poisson time-series generalised additive models using the mgcv package in R to display the concentration response function for each variable (Wood, 2006). All models included the following covariates: (1) influenza epidemics, a binary variable coded as 1 if influenza hospital admission rates were $<$90th percentile and otherwise 0; (2) average temperature (°C) for the previous 3 days (lags 1–3) modelled as a non-linear response with 2 degrees of freedom; (3) average dew point for the previous 3 days (lags 1–3) modelled as a non-linear response with 2 degrees of freedom; (4) same-day average temperature (°C) modelled as a non-linear response with 2 degrees of freedom; and (5) same-day average dew point (°C) modelled as a linear response.

Bushfires and dust storms tend to coincide with high ambient temperatures and higher maximum ozone concentrations, both of which have been independently associated with increased risk of mortality (Anderson and Bell, 2009; Ito et al., 2005). We therefore conducted sensitivity analyses to examine the contribution of temperature and 1 h maximum daily ozone to mortality during severe pollution events. We also examined the sensitivity of our results the cut point used to define a severe pollution event (99th vs 95th percentile).

F. Johnston et al. / Environmental Research 111 (2011) 811–816

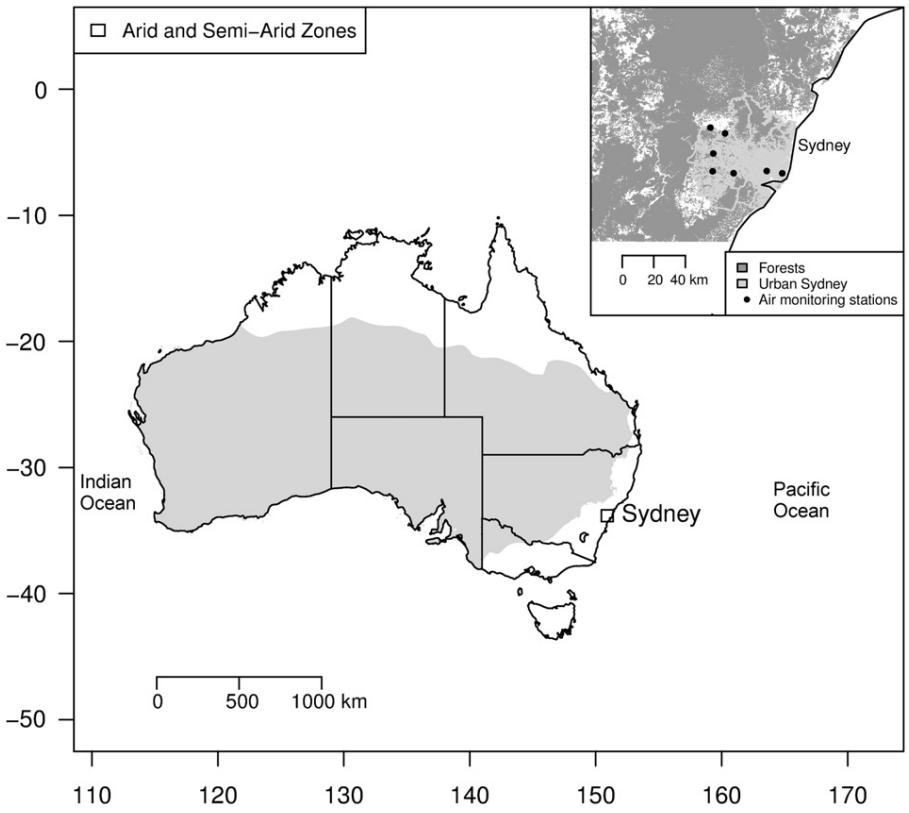


Fig. 1. Map of Australia showing the location of Sydney in relation to arid and semi-arid eco-climatic regions of Australia, the source of most atmospheric dust (based on Hutchinson et al., 2005). Inset shows the urban areas of Sydney, the location of air monitoring stations and adjacent forested areas, the source of most severe bushfire smoke events (Parsons et al., 2006).

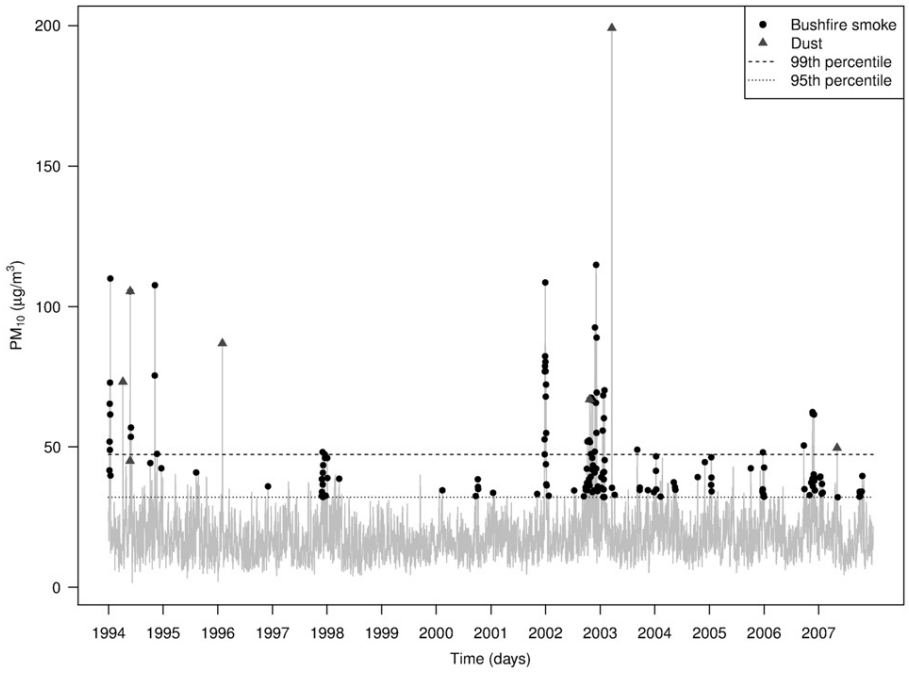


Fig. 2. Daily PM_{10} Sydney, Australia 1994–2007 showing identified bushfire smoke and dust pollution events. The 95th and 99th percentiles are indicated.

3. Results

There were 284,326 deaths during the 13.5 year study period. Nearly half were coded as cardiovascular while approximately 10% were coded as respiratory. There were 52 days on which PM_{10} exceeded the 99th percentile of its distribution (47.3 μg/m^3). All days were successfully identified as being smoke events ($N=46$), dust events ($N=4$), or both ($N=2$). Dust events, though far fewer

in number, tended to have higher particle concentrations. Summary statistics for covariates are listed in Table 1 and results of the analyses are presented in Table 2.

Smoke events were associated with a 5% increase in non-accidental mortality at a lag of one day OR (95%CI) 1.05 (95%CI: 1.00–1.10). When same-day average temperature was excluded

Table 1
Summary statistics for daily pollutant, fire smoke, dust, meteorological and mortality, Sydney, 1994–2007.

Exposure or outcome	Number of observations	Mean	Minimum	First quartile	Median	Third quartile	Maximum
PM_{10} 24 h city wide average ($\mu g/m^3$)	5113	18.3	1.6	12.8	16.9	21.9	199.2
PM_{10} smoke days ($\mu g/m^3$)	48	67.3	47.3	52.2	63.9	75.8	114.8
PM_{10} dust days ($\mu g/m^3$)	6	96.8	49.6	68.4	80.01	100.8	199.2
PM_{10} background days ($\mu g/m^3$)	5061	17.8	1.64	12.8	16.8	21.7	47.3
24hr ave temp (°C), all days	5113	18.3	7.1	14.5	18.3	21.8	33.9
24 h ave Temp (°C), smoke days	48	23.8	12.8	21.0	23.9	26.8	32.8
24 h ave Temp (°C), dust days	6	22.3	17.0	20.6	23.0	24.8	25.5
24 h Dew Point (°C)	5113	10.6	−5.5	6.7	11.0	14.9	21.9
Ozone 1 h max (ppb)	5113	3.2	0.3	2.4	2.9	3.7	12.0
Non-accidental mortality	284,326	58	30	50	57	64	103
Cardiovascular mortality	120, 686	25	7	20	24	29	58
Respiratory mortality	24, 618	5	1	3	5	7	20

Table 2
Associations between non-accidental, cardiovascular, respiratory and combined cardiorespiratory mortality and severe pollution events due to bushfire smoke and dust for the same day (lag 0) and lags of 1–3 days, Sydney 1994–2007. Results of sensitivity analyses for meteorological covariates, same day maximum ozone, and cut point used to define pollution event are included.

Outcome	Model	Lag	BUSHFIRE SMOKE EVENT			DUST EVENT		
			Odds ratio	Lower 95%CI	Upper 95%CI	Odds ratio	Lower 95%CI	Upper 95%CI
Non-accidental mortality	Basic model[a]	0	1.02	0.98	1.05	1.04	0.92	1.18
		1	**1.05**	**1.00**	**1.10**	0.96	0.85	1.09
		2	1.00	0.95	1.04	0.96	0.85	1.09
		3	1.00	0.96	1.05	**1.16**	**1.03**	**1.30**
	Basic model[a] with exclusion of same day temperature	0	**1.05**	**1.00**	**1.09**	1.07	0.94	1.21
		1	**1.05**	**1.00**	**1.09**	0.97	0.86	1.10
		2	0.98	0.94	1.03	0.96	0.85	1.08
		3	1.00	0.95	1.04	**1.15**	**1.02**	**1.29**
	Basic model[a] with inclusion maximum ozone	0	1.02	0.97	1.07	1.04	0.92	1.17
		1	**1.05**	**1.01**	**1.10**	0.96	0.85	1.09
		2	0.99	0.95	1.04	0.90	0.85	1.09
		3	1.00	0.97	1.06	**1.16**	**1.03**	**1.30**
	Basic model[a] with 3 df for all metrological covariates	0	1.02	0.97	1.06	1.12	0.97	1.30
		1	**1.05**	**1.00**	**1.09**	1.03	0.89	1.19
		2	0.99	0.95	1.04	0.96	0.83	1.12
		3	1.01	0.96	1.06	**1.20**	**1.04**	**1.39**
	Basic model[a] 95th percentile used to define a pollution event	0	1.01	0.98	1.04	1.06	0.95	1.19
		1	1.00	0.98	1.04	1.01	0.90	1.13
		2	0.94	0.96	1.01	0.94	0.84	1.05
		3	1.00	0.97	1.02	**1.17**	**1.05**	**1.30**
Cardiovascular mortality	Basic model[a]	0	1.06	0.97	1.17	1.12	0.87	1.44
		1	1.07	0.98	1.18	0.99	0.76	1.30
		2	0.98	0.89	1.08	1.02	0.77	1.34
		3	1.01	0.92	1.11	1.10	0.85	1.42
	Basic model[a] with exclusion of same day temperature	0	**1.10**	**1.00**	**1.20**	1.14	0.89	1.47
		1	1.07	0.97	1.17	0.99	0.76	1.30
		2	0.96	0.88	1.06	1.02	0.77	1.34
		3	1.01	0.92	1.11	1.10	0.85	1.42
Respiratory mortality	Basic model[a]	0	1.00	0.79	1.25	0.89	0.49	1.63
		1	1.09	0.88	1.36	0.55	0.26	1.17
		2	0.96	0.77	1.20	1.02	0.54	1.91
		3	0.94	0.75	1.17	1.23	0.67	2.24
Combined cardiorespiratory mortality	Basic model[a]	0	1.02	0.97	1.07	1.05	0.88	1.24
		1	1.05	0.97	1.15	1.08	0.86	1.37
		2	0.98	0.89	1.02	1.07	0.79	1.31
		3	1.00	0.92	1.08	1.12	0.89	1.42

[a] Basic model covariates: (1) Flu epidemics, a binary variable coded as 1 if influenza hospital admission rates were $<$90th percentile and otherwise 0; (2) average temperature (°C) for the previous 3 days (lags 1–3) modelled as a non-linear response with 2 degrees of freedom; (3) average dew point for the previous 3 days (lags 1–3) modelled as a non-linear response with 2 degrees of freedom; (4) same-day average temperature (°C) modelled as a non-linear response with 2 degrees of freedom; and (5) same-day average dew point (°C) modelled as a linear response. Bold signifies $p \leq 0.05$.

F. Johnston et al. / Environmental Research 111 (2011) 811–816

from the model, associations of a similar magnitude were also observed on the day of the event (lag 0) OR (95%CI) 1.05 (95%CI: 1.00–1.09). As there were fewer dust events, our effect estimates for this exposure were much less precise. We observed a 15% increase in non-accidental mortality three days following the event OR (95%CI) 1.16 (95%CI: 1.03–1.30). In contrast to smoke events, the outcomes for dust events were not sensitive to the inclusion of same-day temperature. Daily maximum ozone and the number of degrees of freedom used for the meteorological covariates did not influence results for smoke or dust events (Table 2). We did not find any associations with mortality when we included less severe smoke (and dust) pollution events in which PM_{10} exceeded the 95th percentile of 32.0 $\mu g/m^3$ (Table 2.) This is not surprising given that almost half of these events were not validated as being attributable to smoke or dust, and were therefore included as background concentrations.

A larger same day association was observed with smoke events and cardiovascular mortality, but only when same day temperature was excluded from the model OR (95%CI) 1.10 (95%CI: 1.00–1.20). Associations between same day dust events also had a larger point estimates for cardiovascular than for non-accidental mortality, but neither association approached statistical significance. Although we looked for associations between pollution events and respiratory mortality, our results were imprecise as deaths attributed to respiratory causes occur less frequently than those attributed to cardiovascular causes (Table 1). In all analyses of respiratory mortality the 95% confidence intervals were very wide and all included the null (Table 2).

4. Discussion

We have described the associations between non-accidental mortality and severe pollution events caused by bushfire smoke and dust. The impacts were greatest on the day-of and day following a smoke event, and three days following a dust event. During smoke events a high same-day temperature contributed to the same-day effect, but did not affect the lagged association with mortality. Cardiovascular mortality had a similar overall pattern to non-accidental mortality with greater effect size estimates for events, although the results did not reach statistical significance.

The strengths of this study include the long time period for which exposure and event data were available and the number of smoke events in the analyses. Many other studies have been limited to single event days or short-term episodes. However, the infrequent nature of dust storms resulted in very few events available for analysis and, consequently, the results were imprecise. Our research was limited by the need to use city-wide rather than individual level exposure assessment and not all regions of the city would have been equally affected by the smoke or dust events or other meteorological exposures.

Other studies on the association between bushfire smoke and mortality are limited, but their findings have been consistent with those reported here. For example, Morgan et al. (2010) found a 0.8% increase in mortality per 10 $\mu g/m^3$ rise in PM_{10} (95%CI= 0.24–1.86%) during 32 bushfire days in Sydney, but reported no association with cardiovascular mortality. Hänninen et al. (2009) and Sastry (2002) both found increases in mortality in the order of 1% per 10 $\mu g/m^3$ rise in PM_{10} associated with forest fires in Finland and Kuala Lumpur, respectively. Our results are broadly consistent with the current evidence concerning urban particulate air pollution which has been causally associated with daily non-accidental and cardiovascular mortality (Pope and Dockery, 2006).

Studies on the association between dust storms and mortality are even more limited than those on bushfire smoke. Perez et al. (2008) studied 90 Saharan dust intrusions in Barcelona over a two year period and observed a large and statistically significant increase in the mortality associated with coarse and fine PM during dust episodes. Mortality on dust days increased 8.4% per 10 $\mu g/m^3$ of particulate matter at a lag of 1 day. In a study of 39 Asian dust events in Taipei, Chen et al. (2004) observed a 5% increase in mortality at a lag of 2 days, but this did not reach statistical significance and their analyses were not adjusted for meteorology, other pollutants or influenza. Finally, Kwon et al. (2002) studied 28 Asian dust events that affected Seoul during 1995–1998. In a time series analysis (adjusted for meteorological variables but not influenza) they also found associations with non-accidental mortality (3.4% increase; 95%CI=0.5–7.4%) at a lag of 2 days.

The most notable difference between the dust and smoke events in our study was in the lag at which the associations were observed. Although the lag 3 association for dust events could be a chance finding (given the small number of events in the study), it is consistent with other available epidemiological evidence. Both Chen and Kwon reported lag times of 2 days until clinical outcomes were observed. Further, in their study of Asian dust events and hospital admissions in Taiwan, Bell and colleagues noted considerable temporal variation (1–3 days) in association within a single clinical outcome, depending upon which indicator of dust had been included in the analyses (Bell et al., 2008). It is plausible that the spectrum of health impacts associated with dust and smoke pollution events could be different from each other as they have markedly different compositions. Dust is derived from crustal particles, has a lower carbon content, a higher proportion of particles of larger size fractions and a large component of biological particles, such as bacteria, viruses and fungi (Griffin, 2007). It is plausible that longer lag times to adverse health outcomes following dust exposure could be explained by human patho-physiological responses to dust-borne micro-organisms. For example, the development of life-threatening sepsis takes longer than the precipitation of cardiac events, a well described outcome associated with smaller combustion-generated particles (Dennekamp et al., 2010). We were not able to explore this hypothesis in our analysis because there were too few daily deaths attributed to either respiratory or septic causes to enable a meaningful analysis.

The threshold for defining a severe pollution event in this study was the 99th percentile of the distribution of PM_{10} over the entire study period. This equated to 47.3 $\mu g/m^3$, just below Australia's 24-h air quality standard for PM_{10} (50 $\mu g/m^3$). However, in comparison to the air quality in many other countries, the pollution events we studied were relatively minor. Even during the most severe smoke event the city-wide average (114.8 $\mu g/m^3$) did not approach the current US air quality standard for PM_{10} (150 $\mu g/m^3$). Therefore, much higher mortality impacts might be evident in regions where large populations are recurrently exposed to very severe episodes of fire smoke, such as those sometimes cased by annual burning in South East Asia (van der Werf et al., 2010). We did not find associations with mortality when we included events in which PM_{10} exceeded the 95th percentile (32 $\mu g/m^3$) as a sensitivity analysis, which is not surprising. While including less severe pollution events provided additional data points for analysis, background pollution sources remain major contributors to ambient particle concentrations between the 95th and 99th percentiles (Johnston et al., 2011).

One limitation of these analyses is that we cannot determine which elements of the smoke and dust events were most important in the observed associations with mortality. Given that events were defined according to PM_{10} concentrations, we assume that particulate matter is responsible for some of the effect. We have also demonstrated that same-day temperature is an important contributor during bushfire events. However, it is possible that population-wide responses such as increased psychological stress also play a role. Similarly, other environmental

exposures (e.g. co-pollutants in bushfire smoke or biological particles associated with dust events) could also contribute to adverse health outcomes (Jalava et al., 2006; Perez et al., 2008).

The long distance transport of dust and smoke and the associated health effects highlight the need to consider the nexus between the health of human populations and our landscapes. A hotter and drier climate favors more fire and drought, with resultant slower recovery of vegetation, and a higher risk of overgrazing from livestock (Mitchell et al., 2010; Wang and Hendon, 2007). Thus, the potential for harms from both smoke and dust events are likely to increase in Australia, and other regions with similar environments, as the global climate changes.

5. Conclusions

In summary, we have described associations between non-accidental mortality and severe pollution events caused by bushfire smoke and dust. During smoke events a high same-day temperature contributed to the same-day but not lagged associations with mortality. The implications of this study are threefold: (1) the need for better evidence to inform public health responses during dust storm and bushfire smoke events as both higher temperatures and pollution are likely to play a role in population health outcomes and need to be addressed as part of the public health response; (2) the need for greater acknowledgement of the linkages between the management of dust-prone rangelands and flammable vegetation with the wellbeing of rural and urban communities; and (3) the urgent need for mitigation of climate change to minimise its many inter-related consequences including severe bushfires, worsening air quality, heat stress and drought.

Acknowledgments

Grant McTainsh, John Leys, Grant Williamson and Talia Portner contributed to the identification of severe pollution episodes.

References

Anderson, B., Bell, M., 2009. Weather-related mortality: how heat, cold, and heat waves affect mortality in the United States. Epidemiology 20 (2), 205.

Atkinson, R.W., Anderson, H.R., Sunyer, J., Ayres, J., Baccini, M., Vonk, J.M., et al., 2001. Acute effects of particulate air pollution on respiratory admissions. American Journal of Respiratory & Critical Care Medicine 164, 1860–1866.

Australian Bureau of Statistics. Causes of Death Australia1964–2007 (3303.0). Health Statistics Unit, Brisbane, GPO Box 9817, 4000.

Bell, M.L., Levy, J.K., Lin, Z., 2008. The effect of sandstorms and air pollution on cause-specific hospital admissions in Taipei, Taiwan. Occupational and Environmental Medicine 65 (2), 104.

Bowman, D.M.J.S., 2000. Australian Rainforests: Islands of Green in a Land of Fire. Cambridge University Press.

Chen, Y.S., Sheen, P.C., Chen, E.R., Liu, Y.K., Wu, T.N., Yang, C.Y., 2004. Effects of Asian dust storm events on daily mortality in Taipei, Taiwan. Environmental Research 95 (2), 151–155.

Confalonieri, U., Menne, B., Akhtar, R., Ebi, K.L., Hauengue, M., Kovats, R.S., et al., 2007. Human health, climate change 2007: impacts, adaptation and vulnerability. In: Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E. (Eds.), Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK, pp. 402.

Dennekamp, M., Akram, M., Abramson, M.J., Tonkin, A., Sim, M.R., Fridman, M., et al., 2010. Outdoor air pollution as a trigger for out-of-hospital cardiac arrests. Epidemiology 21 (4), 494.

Griffin, D.W., 2007. Atmospheric movement of microorganisms in clouds of desert dust and implications for human health. Clinical Microbiology Reviews 20 (3), 459.

Hanigan, I.C., Hall, G., Dear, K.B., 2006. A comparison of methods for calculating population exposure estimates of daily weather for health research. International Journal Health Geography 5, 38.

Hänninen, O., 2008. Challenges in estimating the health effects of biomass smoke. Response to Sverre Vedal and Steven J. Dutton: Wildfire air pollution and daily mortality in a large urban area. Environmental Research 106 (3), 423.

Hänninen, O.O., Salonen, R.O., Koistinen, K., Lanki, T., Barregard, L., Jantunen, M., 2009. Population exposure to fine particles and estimated excess mortality in Finland from an East European wildfire episode. Journal of Exposure Science and Environmental Epidemiology 19 (May/June), 414–422.

Hashizume, M., Ueda, K., Nishiwaki, Y., Michikawa, T., Onozuka, D., 2010. Health effects of Asian dust events: a review of the literature. Nippon Eiseigaku Zasshi (Japanese Journal of Hygiene) 65, 413–421.

Hutchinson, M.F., McIntyre, S., Hobbs, R.J., Stein, J.L., Garnett, S., Kinloch, J., 2005. Integrating a global agro-climatic classification with bioregional boundaries in Australia. Global Ecology and Biogeography 14, 197–212.

Ito, K., De Leon, S.F., Lippmann, M., 2005. Associations between ozone and daily mortality: analysis and meta-analysis. Epidemiology 16 (4), 446.

Jalava, P.I., Salonen, R.O., Hälinen, A.I., Penttinen, P., Pennanen, A.S., Sillanpää, M., et al., 2006. In vitro inflammatory and cytotoxic effects of size-segregated particulate samples collected during long-range transport of wildfire smoke to Helsinki. Toxicology and Applied Pharmacology 215 (3), 341–353.

Jayachandran, S., 2005. Air Quality and Infant Mortality During Indonesia's Massive Wildfires in 1997. Department of Economics UCLA, California, p. 55.

Johnston, F.H., Hanigan, I.C., Henderson, S.B., Morgan, G.G., Portner, T., Williamson, G.J., Bowman, D.M.J.S., 2011. Creating an integrated historical record of extreme bushfire smoke events in Australian cities from 1994 to 2007. Journal of the Air and Waste Management Association 61, 390–398.

Kwon, H.J., Cho, S.H., Chun, Y., Lagarde, F., Pershagen, G., 2002. Effects of the Asian dust events on daily mortality in Seoul, Korea. Environmental Research 90 (1), 1–5.

Maclure, M., 1991. The case-crossover design: a method for studying transient effects on the risk of acute events. American Journal of Epidemiology 133 (2), 144–153.

McTainsh, G.H. 1998. 'The Dustwatch project'. Dust Event Database and Dust Storm Index. Sustainable Agriculture: Assessing Australia's Recent Performance. A Report of the National Collaborative Project on Indicators for Sustainable Agriculture Technical Report 70, pp. 65–72.

Mitchell, R.M., Campbell, S.K., Qin, Y., 2010. Recent increase in aerosol loading over the Australian arid zone. Atmospheric Chemistry and Physics 10, 1689–1699.

Morgan, G., Sheppeard, V., Khalaj, B., Ayyar, A., Lincoln, D., Jalaludin, B., et al., 2010. The effects of bushfire smoke on daily mortality and hospital admissions in Sydney, Australia, 1994 to 2002. Epidemiology 21 (1), 47–55.

NASA, 2010. MODIS Rapid Response System, 〈http://rapidfire.sci.gsfc.nasa.gov/〉.

National Casemix and Classification Centre, 2010. IDC-10 Mapping Tables. Available: 〈http://nccc.uow.edu.au/ICD-10am/ICD-10ammappingtables/index.html〉.

NSW Department of Environment Climate Change and Water, 2010. Air Emissions Inventory for the Greater Metropolitan Region in NSW. [Online December 2010].

Parsons, M., Gavran, M., Davidson, J., 2006. National Forest Inventory. Bureau of Rural Sciences, Canberra.

Perez, L., Tobias, A., Querol, X., Künzli, N., Pey, J., Alastuey, A., et al., 2008. Coarse particles from Saharan dust and daily mortality. Epidemiology 19 (6), 800.

Pope, C.A., Dockery, D.W., 2006. 2006 critical review. Health effects of fine particulate air pollution: lines that connect. Journal of the Air and Waste Management Association 56, 709–742.

R Foundation for Statistical Computing, 2006. R: A Language and Environment for Statistical Computing. Version 2.4.0. Vienna, Austria.

Raupach, M., McTainsh, G.H., Leys, J., 1994. Estimates of dust mass in recent major Australian dust storms. Australian Journal of Soil and Water Conservation 7 (3), 20–24.

Sastry, N., 2002. Forest fires, air pollution, and mortality in southeast Asia. Demography 39 (1), 1–23.

Touloumi, G., Samoli, E., Quenel, P., Paldy, A., Anderson, R.H., Zmirou, D., et al., 2005. Short-term effects of air pollution on total and cardiovascular mortality: the confounding effect of influenza epidemics. Epidemiology 16 (1), 49.

van der Werf, G.R., Randerson, J.T., Giglio, L., Collatz, G.J., Mu, M., Kasibhatla, P.S., et al., 2010. Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). Atmospheric Chemistry and Physics Discussions 10, 16153–16230.

Vedal, S., Dutton, S.J., 2006. Wildfire air pollution and daily mortality in a large urban area. Environmental Research 102 (1), 29–35.

Wang, G., Hendon, H.H., 2007. Sensitivity of Australian rainfall to inter-El Nino variations. Journal of climate 20 16, 4211–4226.

WHO, 2005. International Classification of Diseases Version 10, Second ed World Health Organization, Geneva.

Wood, S.N., 2006. Generalized Additive Models: An Introduction with R. CRC Press.

4.5 Paper 6: The effect on hospital admissions of air pollution from bushfires

Title: Martin, K. L., Hanigan, I. C., Morgan, G. G., Henderson, S. B., and Johnston, F. H. (2013). Air pollution from bushfires and their association with hospital admissions in Sydney, Newcastle and Wollongong, Australia 1994-2007. Australian and New Zealand Journal of Public Health, 37(3), 238-243. doi:10.1111/1753-6405.12065

Stage of publication: Published peer-reviewed paper

Journal Impact Factor: 2

My contribution: In this paper I worked with my co-authors to design the analysis based on the hypothesis proposed by Dr Johnston and Prof. Morgan. I took responsibility for the data processing stage to create the exposure data, cleaned the health outcome data, estimated the influenza epidemic outbreaks, and cleaned the public holiday data. I designed the data analysis code and liaised with Dr Martin who conducted the final data analysis processing. I produced figures and summarised the results. I helped to interpret and write about the results and discussion. I also assisted prepare the text for the methods and results section of the paper.

Approvals required for inclusion of Paper 6 in this thesis:

Name	Notes	Signature/date
Dr. Kara L. Martin		2015-08-24
Mr. Ivan Hanigan		2015-09-28
Prof. Geoffrey G. Morgan		2015-11-15
Dr. Sarah B. Henderson		2015-10-06
Dr. Fay H. Johnston		2015-08-21
Copyright	License Number: 3723600034175 License date: Oct 07, 2015 Licensed Content Publisher: John Wiley and Sons	2015-10-07
Ethics Approval	ANU Human Research Ethics Protocol: 2008/199. Date entered 26/04/2008, variation = 03/06/2011 (extension of study period, addition of more data).	2008-04-26

Environment **Article**

Air pollution from bushfires and their association with hospital admissions in Sydney, Newcastle and Wollongong, Australia 1994–2007

Abstract

Objective: We examined the association between validated bushfire smoke pollution events and hospital admissions in three eastern Australian cities from 1994 to 2007.
Methods: Smoke events were defined as days on which bushfire smoke caused the 24-hour citywide average concentration of airborne particles to exceed the 99th percentile of the daily distribution for the study period. We used a time-stratified case-crossover design to assess the association between smoke events and hospital admissions. Odds ratios (OR) and 95% confidence intervals (CI) were estimated for cardiovascular and respiratory conditions on event days compared with non-event days. Models were adjusted for daily meteorology, influenza epidemics and holidays.
Results: Smoke events occurred on 58 days in Sydney (population: 3,862,000), 33 days in Wollongong (population: 406,000) and 50 days in Newcastle (population: 278,000). In Sydney, events were associated with a 6% (OR=1.06, 95%CI=1.02-1.09) same day increase in respiratory hospital admissions. Same day chronic obstructive pulmonary disease admissions increased 13% (OR=1.13, 95%CI=1.05-1.22) and asthma admissions by 12% (OR=1.12, 95%CI=1.05-1.19). Events were also associated with increased admissions for respiratory conditions in Newcastle and Wollongong.
Conclusions: Smoke events were associated with increased hospital admissions for respiratory but not cardiovascular conditions. Large populations are needed to assess the impacts of brief exposures.
Implications: Public health impacts from bushfire pollution events are likely to increase in association with a warming climate and more frequent severe fire weather.
Key words: vegetation fires, bushfire smoke, hospital admissions, particulate air pollution, respiratory disease, cardiovascular disease

Aust NZ J Public Health. 2013; 37:238-43
doi: 10.1111/1753-6405.12065

Kara L. Martin
Menzies Research Institute Tasmania, University of Tasmania

Ivan C. Hanigan
School of Plant Science, University of Tasmania; National Centre for Epidemiology and Population Health, Australian National University, Australian Capital Territory

Geoffrey G. Morgan
University Centre for Rural Health – North Coast, University of Sydney, New South Wales; Northern New South Wales Local Health District

Sarah B. Henderson
British Colombia Centre for Disease Control, Canada

Fay H. Johnston
Menzies Research Institute Tasmania, University of Tasmania

Bushfires are an ever-increasing problem in Australia. The bushfire season is getting longer and episodes of severe fire weather are becoming more frequent.[1-3] As a result, costs from infrastructure losses and deaths have also been increasing.[4] From 1967 to 1999 insurance claims from bushfires averaged $77 million per year. The bushfires in January 2003 caused more than $300 million in claims, and those from the 2009 Victorian bushfires were estimated to be in excess of $4.4 billion.[3,5] In an attempt to reduce fuel loads and the risk of hazardous fires, prescribed burning on public lands has been recommended by The Royal Commission into the Victorian Bushfires of 2009. However, they acknowledge that prescribed burning is a controversial practice as it is risky and resource intensive, and has a range of adverse effects on local communities, especially with respect to air quality.[3]

A management constraint is the relative lack of good-quality evidence specifically concerning the health impacts of bushfire smoke, and uncertainty about the applicability of findings from the vast literature on urban air pollution, most of which concerns pollution derived from industry and transport, rather than vegetation fires. A growing body of research has consistently shown that elevated concentrations of particulate matter (PM) from vegetation fire smoke are associated with respiratory morbidity, especially related to adult asthma and chronic obstructive pulmonary disease (COPD).[6-8] Fewer studies of mortality have been published, with only two[9,10] of five studies[11-13] reporting a significant association with smoke-related PM. Similarly, studies of associations with cardiovascular disease outcomes are limited and report inconsistent findings.[11,14-19] Dennekamp et al.[20] showed an increased risk of out-of-hospital cardiac arrests on days affected by bushfire smoke in Melbourne, Australia. Another study on peat bog smoke found associations with cardio-pulmonary symptoms and heart

Submitted: July 2012 **Revision requested**: December 2012 **Accepted**: February 2013
Correspondence to: Dr Fay Johnston, Menzies Research Institute, UTAS, Private Bag 23, Hobart, Tasmania 7001; e-mail: Fay.Johnston@utas.edu.au

© 2013 The Authors. ANZJPH © 2013 Public Health Association of Australia

failure in North Carolina, US.[21] However, other studies have been less conclusive. Johnston et al.[16] and Hanigan et al.[15] found no associations with admissions for cardiovascular conditions, with the exception of a significant positive association for Indigenous Australians, a high-risk group. Similarly, Henderson et al.[22] found associations for respiratory physician attendances and hospital admissions, but found no associations for cardiovascular outcomes.

Our aim was to investigate the link between PM related to bushfire smoke (PM_{10} or $PM_{2.5}$) and morbidity as measured by overall hospital admissions, and specifically for respiratory and cardiovascular diseases, in a multi-centre study of three eastern Australian cities intermittently affected by bushfire smoke. These were Sydney, Wollongong and Newcastle, with populations of about 3,862,000; 278,000; and 406,000, respectively.[23] All three cities lie on a coastal lowland plain surrounded by elevated sandstone tablelands largely covered with thick eucalypt forests (Figure 1). The main sources of background PM include motor vehicle and industrial emissions, domestic wood smoke and crustal particles.[24] The majority of fire smoke is derived from prescribed and wild fires in the eucalypt forests in the Blue Mountains to the west of the region.[25] Although fire activity is highly variable, these cities are affected by severe pollution events from bushfires for an average of three days each year.[25]

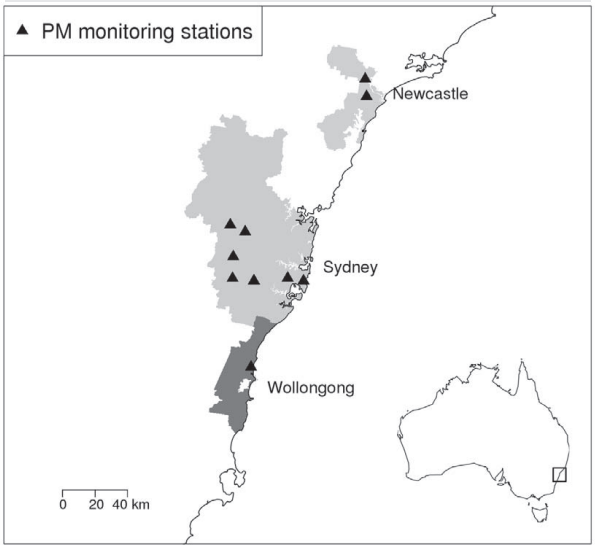
Methods

Exposure measurement

Air quality data were obtained from the NSW Office of Environment and Heritage (www.environment.nsw.gov.au/AQMS/search.htm). They provided 13.5 years of daily average concentrations of PM less than 10 microns (PM_{10}) and less than 2.5 microns ($PM_{2.5}$) in aeronautic diameter. Both PM fractions were measured by tapered element oscillating microbalances at seven monitoring stations in Sydney, one in Wollongong and two in Newcastle (Figure 1). Data were only included if it were available for at least 70% of the days in the time series. We calculated daily city-wide averages of PM_{10} concentrations from 1994 to 2007, and $PM_{2.5}$ concentrations from 1996 to 2007 (1998–2007 in Wollongong). Missing values were imputed using the weighted average values from all proximal monitoring sites in accordance with the protocol from the APHEA (Air Pollution and Health: a European Approach) study.[26]

A database of validated vegetation fire smoke pollution events for this region from 1994-2007 has been previously described.[25] In brief, smoke events were defined as those days when the PM_{10} and/or $PM_{2.5}$ exceeded the 99th percentile of the entire time series and the elevated PM could be attributed to vegetation fires (including bushfires, forestry regeneration and agricultural burning) using media sources, government records and/or remote sensing data. Days above the 99th percentile, but not conclusively validated as being affected by vegetation fires, were classified as background PM days. The 99th percentile was chosen as it corresponds closely to Australian air quality standards and enables a clear delineation between background and bushfire days. Bushfires close to the study populations have occurred since 2007. However the extent to which these influenced city-wide mean concentrations of particulate matter has not been evaluated.

Figure 1: Map of the PM monitoring stations in the cities of Sydney, Wollongong and Newcastle, Australia.



The Australian Bureau of Meteorology provided daily average ambient temperature and humidity, as measured by dew point temperature.[27] The inverse distance weighted averages of all valid weather station observations within 50 kilometres of the population weighted centres of each city were calculated and merged with the health data.[28] Epidemics of influenza were defined as days with hospital admission rates for influenza (ICD9 code 487 and ICD10 codes J10-J11) greater than the 90th percentile of the distribution for greater Sydney metropolitan region.[29] Public and school holidays for the NSW region were collated from a comprehensive listing of events[30] and diaries and calendars.

Outcome data

The study population was defined geographically according to statistical local area of residence. We included all statistical local areas in the Sydney, Newcastle and Wollongong metropolitan areas for which representative air quality monitoring data were available. Hospital admissions data were provided by the Department of Health in NSW, coded according to the World Health Organization's International Classification of Diseases (ICD).[31] Until December 1996, the primary diagnosis for admissions to hospital was coded according to the ICD 9th revision (ICD-9) and thereafter according to the 10th revision (ICD-10). To map from ICD-9 to ICD-10 codes we followed the protocols of the National Casemix and Classification Centre.[32] The association between smoke events and hospital admissions was examined for all non-trauma admissions (ICD9 <800, ICD10 A00-R99), and specifically for respiratory (ICD9 = 460-519; ICD10 = J00-J99 {excluding J95.4-J95.9} R09.1, R09.8) and cardiovascular (ICD9 = 390-459; ICD10 = I00-I99 {excluding I67.3, I68, I88, I97.8, I97.9, I98}, G45 {excluding G45.3}, G46, M30, M31, R58) diagnoses. Respiratory subgroups included asthma (ICD9 = 493; ICD10 = J45-J46); chronic obstructive pulmonary

© 2013 The Authors. ANZJPH © 2013 Public Health Association of Australia

disease (ICD9 = 490-492,494-496; ICD10 = J40-J44, J47, J67); and pneumonia and acute bronchitis (ICD9 = 466,480-486; ICD10 = J12-J17, J18.0, J18.1, J18.8, J18.9, J20, J21). Cardiovascular subgroups included arrhythmia (ICD9 = 427; ICD10 = I46-I49); cardiac failure (ICD9 = 428; ICD10 = I50); cerebrovascular disease (ICD9 = 430-438; ICD10 = I60-I66, I67); and ischemic heart disease (ICD9 = 410-413; ICD10 = I20-I25).

Statistical analyses

The study followed a time-stratified case-crossover design in which the event status (smoke event or background day: a binary variable 1,0) on the day of hospitalisation (or up to three days before) was compared with the event status on control days matched by day of week, month and calendar year. This design controls for the potential effects of day of week, season and long term trends on hospital admissions.[33] We used conditional logistic regression models adjusted for temperature, dew point, influenza epidemics and public holidays to estimate the odds ratio (OR) for hospital admissions associated with smoke events compared with the background. For asthma, we also included school holidays as a covariate in the models.[34] As just 58 bushfire smoke days were available for analysis in Sydney, and slighter fewer in the other two centres, our sample size was limited, and for this reason we did not stratify analyses by gender or age.

To determine the optimal number of degrees of freedom for modelling the meteorological covariates we fit natural cubic splines with varying degrees of freedom for each temperature and humidity variable and assessed the difference in the Akaike Information Criterion (AIC). The models included the covariates as follows: average temperature and humidity (ºC – same day average temperature/humidity {measured as dew point temperature} modelled as a non-linear variable with 3 degrees of freedom, df); lagged temperature and humidity (ºC – average temperature/humidity for the previous 3 days {lags1-3} modelled as a non-linear variable with 3 df); and a binary variable for influenza epidemic (if influenza hospital admissions rates were >90th percentile), public holidays and school holidays (for asthma only). A sensitivity analysis was undertaken to ascertain whether Ozone changed the associations. All analyses were conducted using the R statistical software package.[35]

Ethical approval

This study was approved by the Tasmanian Human Research Ethics Committee (H0010047) and the Human Research Ethics Committee of the Australian National University (2008/199).

Results

During the study period there were 58 days validated smoke event days in Sydney, 33 in Wollongong and 50 in Newcastle. The vast majority were due to bushfires, rather than agricultural or other types of planned vegetation fires. Summary statistics for daily PM_{10} and $PM_{2.5}$ and meteorological variables for each city are shown in Table 1. In Sydney, there were a total of 3,141,017 non-trauma hospital admissions, 273,034 in Wollongong and 345,736 in Newcastle. Of those, between 17% and 19% were coded as cardiovascular admissions and between 13% and 15% as respiratory admissions, depending on the city.

In Sydney, smoke events were significantly (p<0.05) associated with a 2% increase in all non-trauma hospital admissions on the same day as the event (OR=1.02, 95%CI=1.00–1.03). Smoke events were not associated with the total number of same day hospital admissions in Wollongong or Newcastle. These results were all independent of temperature (same day and previous three days), humidity (same day and previous three days), influenza epidemics, and public holidays.

The associations between smoke events and hospital admissions for all respiratory conditions are shown in Figure 2 and associations

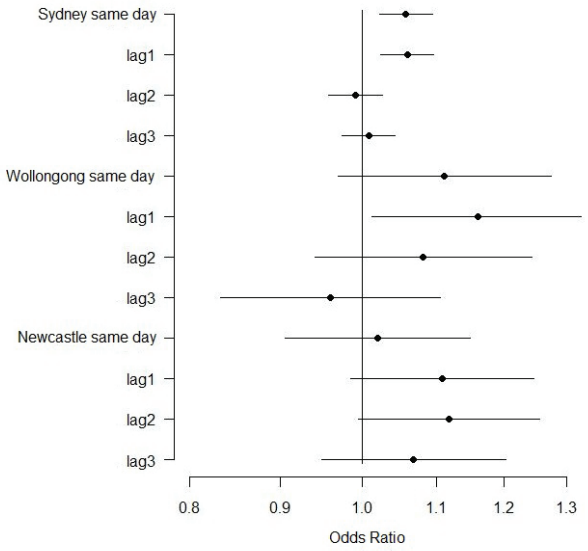
Table 1: Summary statistics for particulate concentration-meteorological variables and hospital admissions – Greater NSW – 1994-2007.

	Sydney			Wollongong			Newcastle		
	N	Mean	Range	N	Mean	Range	N	Mean	Range
PM10 24hr city-wide average µg/m^3	4,929	18.4	1.6-199.2	4,624	19.0	2.5-280.5	4,800	19.4	2.0-160.9
PM2.5 24hr city-wide average µg/m^3	4,006	10.2	2.1-100.2	3,359	9.4	1.6-112.2	3,868	9.7	1.8-61.9
PM10 smoke days µg/m^3*	48	67.3	47.3-114.8	31	68.8	50.9-280.5	34	67.1	49.9-160.9
PM2.5 smoke days µg/m^3 *	36	43.9	27.4-100.2	31	38.3	24.7-112.2	37	34.5	25.2-61.9
24hr avg Temp ˚C-all days	4,929	18.3	7.2-33.9	4,929	17.5	8.0-32.4	4,929	18.3	8.7-33.1
24hr Dew Point ˚C	4,929	10.7	–5.5-21.9	4,929	11.4	–6.2-23.5	4,929	12.5	–3.7-23.7
Hospital Admissions:									
Total Admissions	3,141,017	2770	1690-4893	273,034	241	78-421	345,736	305	117-515
Cardiovascular admissions	519,621	105	14-169	51,833	46	15-85	60,522	56	8-104
Arrhythmia	72,198	15	1-38	7,456	2	0-9	7,577	2	0-11
Cardiac Failure	71,352	14	1-39	6,739	1	0-8	9,530	2	0-9
Cerebrovascular Disease	89,819	18	2-36	9,563	2	0-9	11,205	2	0-9
Ischemic Heart Disease	180,852	37	5-66	18,824	4	0-15	24,369	5	0-19
Respiratory admissions	483,028	98	29-215	35,796	32	6-83	46,394	41	11-93
Asthma	76,668	21	1-87	4,823	1	0-10	5,704	2	0-9
COPD†	89,030	18	2-55	9,141	2	0-11	10,021	2	0-11
Pneumonia and Acute Bronchitis	142,634	29	1-87	10,530	2	0-12	14,607	3	0-16

*N – number of observations; * - Total number of days in excess of PM10 and/or PM2.5 = 58; † - Chronic Obstructive Pulmonary Disease*

© 2013 The Authors. ANZJPH © 2013 Public Health Association of Australia

Figure 2: Associations between respiratory hospital admissions and pollution events from bushfire smoke for the same day and at lags of 1-3 days in Sydney, Wollongong and Newcastle, 1994-2007.



with specific diagnoses are listed in Table 2. In Sydney, associations with respiratory outcomes of similar magnitude were generally observed on the day of a smoke event and the day after a smoke event. The same day increase was 5% (OR=1.05, 95%CI=1.02-1.09) for all respiratory admissions, 12% (OR=1.12, 95%CI=1.05-1.19) for asthma admissions, and 13% (OR=1.13, 95%CI=1.05-1.22) for COPD admissions. In the other cities with much smaller populations, associations with all respiratory admissions were more variable and tended to be greatest on the day after the smoke event. In the smaller cities, there was a trend towards a lagged association with pneumonia and bronchitis, which was statistically significant in Newcastle (Figure 2). Associations with specific respiratory diagnoses, although they tended to be positive, were less consistent and lacked precision (Table 2).

Smoke events were not associated with cardiovascular hospital admissions (data not shown) nor any subgroup of cardiovascular disease in any city (Table 3). None of these results were sensitive to the inclusion of Ozone, a pollutant known to have potential health impacts.

Discussion

We found that smoke events were associated with increased hospital admissions for respiratory but not cardiovascular diseases. Our findings are consistent with previous work reporting associations with respiratory outcomes, particularly chronic lower respiratory conditions such as asthma and COPD.[20] We found similar results in all three cities, although the results for the two smaller cities were less precise than those for Sydney. Our findings highlight the difficulty of studying population level impacts of exposure to landscape fire smoke. Despite a long time series, the population centres of fewer than 500,000 people did not provide enough power

Table 2: Associations between respiratory hospital admissions and severe pollution events due to bushfire smoke, for the same day (lag 0) and lags of 1 to 3 days, in greater Sydney 1994-2007.

Outcome	City	Lag	Odds ratio	95% Confidence Interval (Lower, Upper)
Asthma				
	Sydney	0	1.12	(1.05, 1.19)
		1	1.10	(1.04, 1.17)
		2	1.12	(1.05, 1.19)
		3	1.06	(0.99, 1.12)
	Newcastle	0	0.97	(0.75, 1.26)
		1	1.02	(0.80, 1.29)
		2	1.06	(0.84, 1.35)
		3	1.10	(0.88, 1.37)
	Wollongong	0	1.03	(0.80, 1.33)
		1	1.04	(0.82, 1.33)
		2	1.07	(0.83, 1.38)
		3	1.06	(0.83, 1.34)
Chronic Obstructive Pulmonary Disease				
	Sydney	0	1.13	(1.05, 1.22)
		1	1.12	(1.04, 1.21)
		2	1.02	(0.95, 1.10)
		3	1.05	(0.97, 1.13)
	Newcastle	0	1.18	(0.94, 1.48)
		1	1.20	(0.96, 1.51)
		2	1.09	(0.86, 1.37)
		3	1.35	(1.08, 1.69)
	Wollongong	0	1.25	(0.98, 1.60)
		1	1.09	(0.84, 1.40)
		2	1.31	(1.02, 1.67)
		3	1.01	(0.78, 1.29)
Pneumonia and Acute Bronchitis				
	Sydney	0	1.05	(0.98, 1.12)
		1	1.05	(0.98, 1.12)
		2	0.98	(0.92, 1.05)
		3	1.00	(0.94, 1.07)
	Newcastle	0	1.04	(0.84, 1.27)
		1	1.24	(1.01, 1.52)
		2	1.26	(1.03, 1.55)
		3	1.02	(0.82, 1.26)
	Wollongong	0	1.07	(0.81, 1.40)
		1	1.28	(0.97, 1.68)
		2	1.04	(0.78, 1.40)
		3	1.07	(0.80, 1.42)

Model covariates: (1) same-day average temperature (ºC) modelled as a non-linear response with 3 degrees of freedom; (2) same-day average humidity (as dew point) (ºC) modelled as a non-linear response with 3 degrees of freedom; (3) average temperature (ºC) for the previous 3 days (lags1-3) modelled as a non-linear response with 3 degrees of freedom; (4) average dew point for the previous 3 days (lags1-3) modelled as a non-linear response with 3 degrees of freedom; (5) influenza epidemics, a binary variable coded as 1 if influenza hospital admission rates were < 90th percentile and otherwise 0; and (6) public holidays, a binary variable coded as 1 or 0.

Additionally for asthma: school holidays were also modelled as a binary variable (1,0).

Bold p≤0.05

© 2013 The Authors. ANZJPH © 2013 Public Health Association of Australia

Table 3: Associations between cardiovascular hospital admissions and severe pollution events due to bushfire smoke, for the same day (lag 0) and lags of 1 to 3 days, in three NSW cities 1994-2007.

				95% Confidence Interval	
Outcome	**City**	**Lag**	**Odds ratio**	**Lower**	**Upper**
Arrhythmia					
	Sydney	0	0.95	0.88	1.03
		1	0.93	0.86	1.01
		2	0.96	0.88	1.04
		3	0.95	0.88	1.03
	Newcastle	0	1.11	0.84	1.47
		1	0.91	0.68	1.21
		2	0.86	0.64	1.14
		3	1.06	0.79	1.41
	Wollongong	0	0.99	0.92	1.07
		1	1.01	0.94	1.09
		2	0.97	0.90	1.04
		3	1.02	0.95	1.10
Cardiac Failure					
	Sydney	0	1.01	0.93	1.11
		1	0.92	0.84	1.00
		2	0.98	0.90	1.07
		3	1.05	0.96	1.14
	Newcastle	0	1.03	0.78	1.36
		1	1.08	0.82	1.42
		2	1.06	0.81	1.39
		3	1.06	0.81	1.39
	Wollongong	0	0.79	0.57	1.09
		1	0.98	0.70	1.35
		2	1.01	0.72	1.42
		3	0.86	0.62	1.20
Cerebrovascular Disease					
	Sydney	0	0.99	0.92	1.07
		1	1.01	0.94	1.09
		2	0.97	0.90	1.04
		3	1.02	0.95	1.10
	Newcastle	0	0.93	0.74	1.17
		1	0.96	0.77	1.21
		2	0.90	0.71	1.14
		3	1.02	0.82	1.28
	Wollongong	0	0.92	0.71	1.20
		1	0.99	0.77	1.27
		2	1.09	0.84	1.40
		3	1.08	0.84	1.40
Ischemic Heart Disease					
	Sydney	0	0.99	0.94	1.05
		1	1.02	0.97	1.08
		2	1.03	0.98	1.08
		3	0.98	0.93	1.03

to reliably detect associations with well-established respiratory outcomes when compared with Sydney, which has a population of approximately 4 million. We did not find any association with cardiovascular admissions, even in the large population centre of Sydney. One explanation is that there is truly no association or it could reflect the heterogeneity of the underlying causes of each outcome we examined. However, several lines of evidence suggest that adverse cardiovascular outcomes from exposure to vegetation fire smoke are probable. For example: (1) Acute exposures to elevations in urban PM have been associated with cardiovascular morbidity and mortality[36] and toxicological studies suggest that smoke particles elicit similar effects to those of urban particles;[37,38] (2) Associations with bushfire smoke and out-of-hospital cardiac arrests has been documented;[20] (3) Cardiovascular admissions have been associated with PM in cities in which the PM is largely derived from biomass combustion;[39,40] and (4) An intervention to reduce levels of indoor wood smoke showed significant improvements in risk factors for cardiovascular disease.[41]

However, despite this fragmentary evidence, just a handful of epidemiological studies examining cardiovascular disease outcomes and fire smoke have been published to date. Most have had either null or inconclusive results[11,14,16,18,22] with the notable exception of a recent study of exposure to peat fire smoke.[21] This could be because health outcomes associated with exposure to fire smoke are challenging to study. The episodic nature of large events makes it difficult to know precisely when and where air quality will be affected and many affected areas do not have routine air quality monitoring programs.[42] Occasionally, large fires affect urban areas where fixed site air quality monitoring networks measure ambient concentrations in conjunction with those from all other sources. However, as discussed above, the peak exposures may be too short-lived to detect all but the most sensitive health outcomes with adequate statistical power.[43] Although we were able to study a large population and relatively large number of vegetation fire

Table 3 cont. : Associations between cardiovascular hospital admissions and severe pollution events due to bushfire smoke, for the same day (lag 0) and lags of 1 to 3 days, in three NSW cities 1994-2007.

Newcastle	0	1.02	0.87	1.19
	1	0.93	0.79	1.10
	2	0.95	0.81	1.12
	3	1.02	0.87	1.20
Wollongong	0	0.97	0.81	1.16
	1	0.92	0.76	1.10
	2	1.12	0.94	1.34
	3	0.97	0.81	1.17

Model covariates: (1) same-day average temperature (ºC) modelled as a non-linear response with 3 degrees of freedom; (2) same-day average humidity (as dew point) (ºC) modelled as a non-linear response with 3 degrees of freedom; (3) average temperature (ºC) for the previous 3 days (lags1-3) modelled as a non-linear response with 3 degrees of freedom; (4) average dew point for the previous 3 days (lags1-3) modelled as a non-linear response with 3 degrees of freedom; (5) influenza epidemics, a binary variable coded as 1 if influenza hospital admission rates were < 90th percentile and otherwise 0; and (6) public holidays, a binary variable coded as 1 or 0.

Additionally for asthma: school holidays were also modelled as a binary variable (1,0).

Bold p≤0.05

© 2013 The Authors. ANZJPH © 2013 Public Health Association of Australia

smoke events, our results generally had wide confidence intervals and lacked precision.

Moreover, some degree of exposure misclassification is inevitable and this would likely have biased our results towards the null. Other possible explanations for our null results for cardiovascular diseases include there being a true difference in the toxicology of vegetation fire smoke compared with background urban air pollution, with a smaller or absent association with cardiovascular diseases. Further, brief exposures of just one or two days as commonly experienced by southern Australian populations could be less hazardous than chronic exposures. Another potential limitation, as with all studies of this type, is the use of fixed site monitors. Our aim was to evaluate the impact of severe city-wide smoke events. However, it is possible that on some occasions smaller smoke plumes were not detected by available monitors and these would not have been classified as severe smoke events.

Bushfires are a major population hazard in Australia and monitoring of PM levels should be available in fire-prone areas, especially in regional and remote areas which are currently unmonitored.

Acknowledgements

This project was supported by grants from the Australian Research Council (LP LP0882048) and the National Health and Medical Research Council (490057).

References

1. Bowman DMJS, Balch JK, Artaxo P, Bond WJ, Carlson JM, Cochrane MA, et al. Fire in the Earth system. *Science.* 2009;324(5926):481-84.
2. Flannigan MD, Krawchuk MA, De Groot WJ, Wotton BM, Gowman LM. Implications of changing climate for global wildland fire. *Int J Wildland Fire.* 2009;18(5):483-507.
3. Teague BG, McLeod RN, Pascoe SM. *The 2009 Victorian Bushfires Royal Commission. Final Report*. Melbourne (AUST): Victorian Bushfires Royal Commission; 2010.
4. Emergency Management Australia. *Disasters Database*. Canberra (AUST): Attorney-General's Department; 2010.
5. Middelmann MH, Australia G. *Natural Hazards in Australia: Identifying Risk Analysis Requirements.* Canberra (AUST): Geoscience Australia; 2007.
6. Naeher LP, Brauer M, Lipsett M, Zelikoff JT, Simpson CD, Koenig JQ, et al. Woodsmoke health effects: a review. *Inhal Toxicol.* 2007;19(1):67-106.
7. Dennekamp M, Abramson MJ. The effects of bushfire smoke on respiratory health. *Respirology.* 2011;16(2):198-209.
8. Henderson SB, Johnston FH. Measures of forest fire smoke exposure and their associations with respiratory health outcomes. *Curr Opin Allergy Clin Immunol.* 2012;12(3):221.
9. Johnston FH, Hanigan IC, Henderson SB, Morgan GG, Bowman DMJS. Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994-2007. *Environ Res.* 2011;111:811-16.
10. Sastry N. Forest fires, air pollution, and mortality in southeast Asia. *Demography.* 2002;39(1):1-23.
11. Morgan G, Sheppeard V, Khalaj B, Ayyar A, Lincoln D, Jalaludin B, et al. The effects of bushfire smoke on daily mortality and hospital admissions in Sydney, Australia, 1994 to 2002. *Epidemiology.* 2010;21(1):47-55.
12. Hanninen OO, Salonen RO, Koistinen K, Lanki T, Barregard L, Jantunen M. Population exposure to fine particles and estimated excess mortality in Finland from an East European wildfire episode. *J Expos Sci Environ Epidemiol.* 2008.
13. Vedal S, Dutton SJ. Wildfire air pollution and daily mortality in a large urban area. *Environ Res.* 2006 102(1):29-35.
14. Delfino RJ, Brummel S, Wu J, Stern H, Ostro B, Lipsett M, et al. The relationship of respiratory and cardiovascular hospital admissions to the southern California wildfires of 2003. *Br Med J.* 2009;66(3):189.
15. Hanigan IC, Johnston FH, Morgan GG. Vegetation fire smoke, indigenous status and cardio-respiratory hospital admissions in Darwin, Australia, 1996-2005: A Time-series Study. *Environ Health.* 2008;7;42.
16. Johnston FH, Bailie R, Pilotto L, Hanigan I. Ambient biomass smoke and cardio-respiratory hospital admissions. *BMC Public Health.* 2007;7:240.
17. Moore D, Copes R, Fisk R, Joy R, Chan K, Brauer M. Population health effects of air quality changes due to forest fires in British Columbia in 2003: estimates from physician-visit billing data. *Can J Public Health.* 2006;97(2):105-8.
18. Mott JA, Mannino DM, Alverson CJ, Kiyu A, Hashim J, Lee T, et al. Cardiorespiratory hospitalizations associated with smoke exposure during the 1997, Southeast Asian forest fires. *Int J Hyg Environ Health.* 2005;208(1-2):75-85.
19. Mott JA, Meyer P, Mannino D, Redd SC, Smith EM, Gotway-Crawford C, et al. Wildland forest fire smoke: health effects and intervention evaluation, Hoopa, California, 1999. *West J Med.* 2002;176(3):162-3.
20. Dennekamp M, Erbas B, Sim M, Glass D, Keywood M, Abramson M, et al. Air pollution from bushfires and out of hospital cardiac arrests in Melbourne Australia. *Epidemiology.* 2011;22(1):S53.
21. Rappold AG, Stone SL, Cascio WE, Neas LM, Kilaru VJ, Carraway MS, et al. Peat Bog Wildfire smoke exposure in rural North Carolina is associated with cardio-pulmonary emergency department visits assessed through syndromic surveillance. *Environ Health Perspect;* 2011;119(10):1415-20.
22. Henderson S, Brauer M, Macnab Y, Kennedy S. Three measures of forest fire smoke exposure and their associations with respiratory and cardiovascular health outcomes in a population-based cohort. *Environ Health Perspect;* 2011;119(9):1266-71.
23. Australian Bureau of Statistics. *3235.0 - Population by Age and Sex, Australia 2006.*Canberra (AUST): ABS; 2007.
24. NSW Environment and Heritage. *Air Emissions Inventory for the Greater Metropolitan Region in NSW*. Sydney (AUST): Government of New South Wales; 2010.
25. Johnston FH, Hanigan IC, Henderson SB, Morgan GG, Portner T, Williamson GJ, et al. Creating an integrated historical record of extreme bushfire smoke events in Australian cities from 1994 to 2007. *J Air Waste Manag Assoc.* 2011;61:390-8.
26. Atkinson RW, Ross A. Acute effects of particulate air pollution on respiratory admissions. Results from APHEA 2 project. *Am J Respir Crit Care Med.* 2001;164(10):1860.
27. National Climate Centre. *Daily or Three Hourly Weather Data for Bureau of Meteorology Stations*. Melbourne (AUST): Bureau of Meteorology;2013.
28. Hanigan I, Hall G, Dear K. A comparison of methods for calculating population exposure estimates of daily weather for health research. *Int J Health Geogr.* 2006;5(1):38.
29. Touloumi G, Samoli E, Quenel P, Paldy A, Anderson RH, Zmirou D, et al. Short-term effects of air pollution on total and cardiovascular mortality: the confounding effect of influenza epidemics. *Epidemiology.* 2005;16(1):49.
30. Thorsen S. timeanddate.com. Stavanger (NOR): Time and Date AS; 1995-2009.
31. World Health Organization. *International Classification of Diseases.* Version 10. 2nd ed. Geneva (CHE): WHO; 2005.
32. National Casemix and Classification Centre. IDC-10 Mapping Tables. Wollongong: University of Wollongong, 2010.
33. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol.* 1991;133(2):144.
34. Lincoln D, Morgan G, Sheppeard V, Jalaludin B, Corbett S, Beard J. Childhood asthma and return to school in Sydney, Australia. *Public Health.* 2006;120(9):854-62.
35. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna (AUT): R foundation for Statistical Computing; 2006.
36. Pope CA, Dockery DW. 2006 Critical Review. Health effects of fine particulate air pollution: Lines that connect. *J Air Waste Manag Assoc.* 2006;56:709-42.
37. Zelikoff JT, Chen LC, Cohen MD, Schlesinger RB. The toxicology of inhaled woodsmoke. *J Toxicol Environ Health Part B Crit Rev.* 2002;5(3):269-82.
38. Barregard L, Sällsten G, Gustafson P, Andersson L, Johansson L, Basu S, et al. Experimental exposure to wood-smoke particles in healthy humans: effects on markers of inflammation, coagulation, and lipid peroxidation. *Inhal Toxicol.* 2006;18(11):845-53.
39. Sanhueza PA, Torreblanca MA, Diaz-Robles LA, Schiappacasse LN, Silva MP, Astete TD. Particulate Air Pollution and Health Effects for Cardiovascular and Respiratory Causes in Temuco, Chile: A Wood-Smoke-Polluted Urban Area. *J Air Waste Manag Assoc.* 2009;59(12):1481-8.
40. McGowan J, Hider P, Chacko E, Town G. Particulate air pollution and hospital admissions in Christchurch, New Zealand. *Aust N Z J Public Health.* 2002;26(1):23-9.
41. Allen R, Aspinall T, Carlsten C, Karlen B, Leckie S, Mattson A, et al. The Impact of Portable Air Filters on Indoor Air Pollution and Microvascular Function in a Woodsmoke-Impacted Community. *Epidemiology.* 2009;20(6):S51.
42. Johnston F, Williamson G, Bowman D. A Review of Approaches to Monitoring Smoke from Vegetation Fires for Public Health. *Air Qual Climate Change.* 2010;44(2):17.
43. Hänninen O. Challenges in estimating the health effects of biomass smoke—Response to Sverre Vedal and Steven J. Dutton: Wildfire air pollution and daily mortality in a large urban area. *Environ Res.* 2008;106(3):423-4.

© 2013 The Authors. ANZJPH © 2013 Public Health Association of Australia

4.6 Paper 7: The effect on mortality and morbidity of heatwaves

Title: Wilson, L. A., Morgan, G., Hanigan, I. C., Johnston, F., Abu-Rayya, H., Broome, R., Gaskin, C., and Jalaludin, B. (2013). The impact of heat on mortality and morbidity in the Greater Metropolitan Sydney Region: a case crossover analysis. Environmental Health : A Global Access Science Source, 12(1), 98. doi:10.1186/1476-069X-12-98

Stage of publication: Published peer-reviewed paper

Journal Impact Factor: 3.4

My contribution: I developed the analytical methods and gave oversight and direction of the data analysis conducted by my co-authors. I made major contributions to this paper as the lead data analyst and also contributed editorial comment on interpretations of results.

Approvals required for inclusion of Paper 7 in this thesis:

Name	Notes	Signature/date
Dr. Leigh A. Wilson		2015-09-29
Prof. Geoffrey G. Morgan		2015-11-15
Mr. Ivan Hanigan		2015-09-28
Dr. Fay H. Johnston		2015-08-21
Dr. Hisham Abu-Rayya		Could not be contacted 2015-12-05
Dr. Richard Broome		2015-09-30
Mr. Clive Gaskin		Could not be contacted 2015-12-05
Prof. Bin Jalaludin		2015-09-30
Copyright	Copyright 2013 Wilson et al.; licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.	2015-10-07
Ethics Approval	ANU Human Research Ethics Protocol: 2008/199. Date entered 26/04/2008, variation = 03/06/2011 (extension of study period, addition of more data).	2008-04-26

Wilson et al. Environmental Health 2013, 12:98
http://www.ehjournal.net/content/12/1/98

ENVIRONMENTAL HEALTH

RESEARCH Open Access

The impact of heat on mortality and morbidity in the Greater Metropolitan Sydney Region: a case crossover analysis

Leigh Ann Wilson[1,2], Geoffrey Gerard Morgan[3,4*], Ivan Charles Hanigan[5], Fay H Johnston[6], Hisham Abu-Rayya[7], Richard Broome[8], Clive Gaskin[9] and Bin Jalaludin[10,11]

Abstract

Background: This study examined the association between unusually high temperature and daily mortality (1997–2007) and hospital admissions (1997–2010) in the Sydney Greater Metropolitan Region (GMR) to assist in the development of targeted health programs designed to minimise the public health impact of extreme heat.

Methods: Sydney GMR was categorized into five climate zones. Heat-events were defined as severe or extreme. Using a time-stratified case-crossover design with a conditional logistic regression model we adjusted for influenza epidemics, public holidays, and climate zone. Odds ratios (OR) and 95% confidence intervals were estimated for associations between daily mortality and hospital admissions with heat-event days compared to non-heat event days for single and three day heat-events.

Results: All-cause mortality overall had similar magnitude associations with single day and three day extreme and severe events as did all cardiovascular mortality. Respiratory mortality was associated with single day and three day severe events (95thpercentile, lag0: OR = 1.14; 95%CI: 1.04 to 1.24). Diabetes mortality had similar magnitude associations with single day and three day severe events (95thpercentile, lag0: OR = 1.22; 95%CI: 1.03 to 1.46) but was not associated with extreme events. Hospital admissions for heat related injuries, dehydration, and other fluid disorders were associated with single day and three day extreme and severe events. Contrary to our findings for mortality, we found inconsistent and sometimes inverse associations for extreme and severe events with cardiovascular disease and respiratory disease hospital admissions. Controlling for air pollutants did not influence the mortality associations but reduced the magnitude of the associations with hospital admissions particularly for ozone and respiratory disease.

Conclusions: Single and three day events of unusually high temperatures in Sydney are associated with similar magnitude increases in mortality and hospital admissions. The trend towards an inverse association between cardio-vascular admissions and heat-events and the strong positive association between cardio-vascular mortality and heat-events suggests these events may lead to a rapid deterioration in persons with existing cardio-vascular disease resulting in death. To reduce the adverse effects of high temperatures over multiple days, and less extreme but more frequent temperatures over single days, targeted public health messages are critical.

Keywords: Climate change, Heat-wave, Heat-related illness, Heat threshold, Case-crossover design

* Correspondence: geoff.morgan@ncahs.health.nsw.gov.au
[3]University Centre for Rural Health – North Coast, University of Sydney, Sydney, Australia
[4]North Coast Public Health Unit, Mid North Coast Local Health District, New South Wales, Australia
Full list of author information is available at the end of the article

BioMed Central

© 2013 Wilson et al.; licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Extreme heat events are known to contribute to an increase in mortality and morbidity, particularly in vulnerable populations [1,2]. Early Heat Health Warning Systems (HHWS) have been shown to reduce morbidity and mortality during heat-waves (periods of unusually high temperature) [3-6]. HHWS rely on a number of components including timely and reliable meteorological data, an understanding of who is at increased health risk due to heat-waves and effective public health risk communication [7]. There is mounting evidence that heat-waves are increasing in frequency, duration and intensity as part of climate change [1]. Studies indicate that populations acclimatise to their regional climate and that a population's response to a heat-wave is influenced by this acclimatisation [8]. Understanding the impact of climate on population health is the key to reducing the burden of disease due to climate change in Australia [9].

Much of the research linking heat-waves to excess mortality and morbidity has been conducted in the regions of the northern hemisphere where heat-waves are rare events [10-12], although some northern hemisphere locations do experience heatwaves on a regular basis [13-15]. Although heat-related mortality [9,16-19] and morbidity [9,18-21] studies have been conducted in Australia, study methods and the population groups assessed have varied considerably. This study builds on recent research conducted by Khalaj and colleagues [9] in the Sydney Greater Metropolitan Region (GMR) which used a case-only design to identify underlying conditions that modify the health risk associated with heatwaves. Khalaj and colleagues [9] found that people admitted to hospital with underlying mental and behavioural disorders, diseases of the nervous and circulatory system, diseases of the respiratory system and renal disease are more susceptible to extreme heat events. The case-only study design used by Khalaj is useful in determining which groups of patients are more susceptible to extreme heat events but cannot be used to provide quantitative estimates of the overall increase in deaths or the demand for health services during such events.

This study aims to quantify the association between unusual heat events and mortality as well as morbidity (hospital admission) in metropolitan New South Wales (NSW), Australia and inform the development of a statewide HHWS. In addition, this study will help determine the extent to which northern hemisphere studies are generalisable to an Australian setting where the temperature threshold of unusual heat events is substantially higher than in the northern hemisphere [9]. In Australia there is lack of consensus on the definition of unusual heat events with the Australian Bureau of Meteorology defining a heat-wave as a period of more than a few days in a row of above average temperatures for a given location [22].

Methods

Data

Mortality data for all deaths of NSW residents between 1st July 1997 and 31st December 2007 (inclusive) were obtained from the Australian Bureau of Statistics (ABS) [23]. Data on all hospital admissions of NSW residents from 1st July 1997 to 30th June 2010 inclusive were obtained from the NSW Department of Health [24]. Meteorological data on temperature (daily minimum, maximum, mean), humidity (daily minimum, maximum, mean) and wind speed (at 9 am and 3 pm) were obtained from the Australian Bureau of Meteorology (BoM) for all monitoring stations in the Sydney Greater Metropolitan Area (GMR) from 1997 to 2010.

Climate zones

Our analysis focused on the (southern hemisphere) spring and summer months of September to February. We included spring months because early heat-waves can occur during this time and poor acclimatisation to warm weather following the winter months has the potential to increase the likelihood of heat-related illness [8]. Sydney experiences cooler nights in the inland (Sydney west) zone, and cooler days in the coastal (Sydney east) zone. Previous work by Khalaj et al. [9] used data from all Sydney weather stations to define two Sydney climate zones (Sydney east, Sydney west) by maximising the ratio of between-region to within-region temperature variance. In this study Statistical Local Area (SLA) of residence was used to assign mortality and hospital admissions data to each of the five climate zones of Sydney east, Sydney west, Gosford/Wyong, Newcastle, Illawarra [25]. Daily weather data were obtained for all BoM stations within 50 km of the population-weighted centre of each SLA [26]. The daily average was calculated as the inverse distance weighted average of all validated observations on each day. The distance weighted averages from the population centres were used because we assumed these estimates more closely relate to the experience of the population, rather than that at airports or other locations [27]. The population centre was calculated using data from the 2006 Australian Census [28]. Figure 1 illustrates the study region including the temperature zones, weather and air pollution monitoring stations.

Health outcome measures

Hospital admission and mortality data based upon study subjects' usual place of residence were mapped to SLA 2006 boundaries. While our literature review identified many variations in disease categorisations for both mortality and hospital admissions in previous studies, these variations were generally minor and concerned relatively infrequent diagnoses.

Wilson et al. Environmental Health 2013, 12:98
http://www.ehjournal.net/content/12/1/98

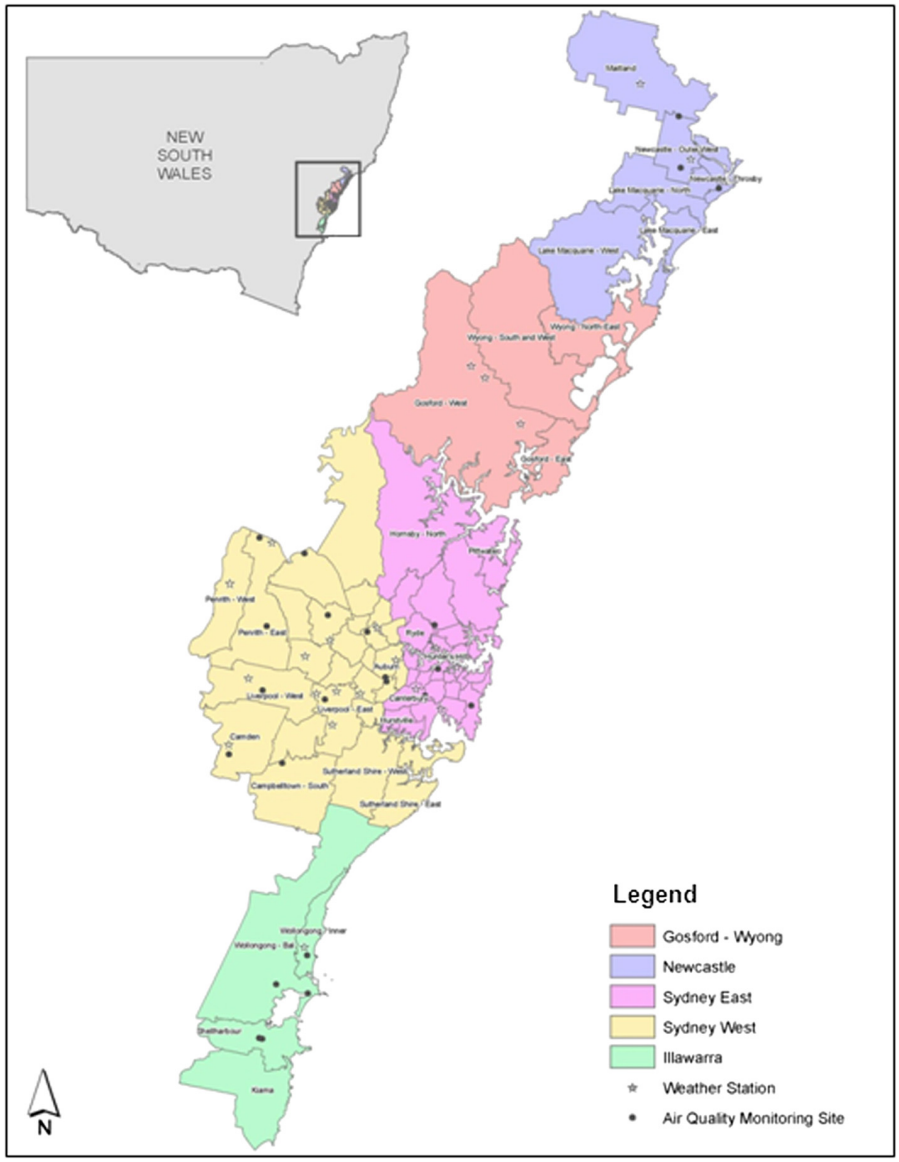


Figure 1 Map of study region including temperature zones and weather and air pollution monitoring stations.

Mortality was grouped into categories based on the underlying cause of death (i.e. the disease or injury which initiated the train of events leading directly to death). Categories used were based on International Classification of Diseases – 10 (ICD-10) coding informed by previous research (Additional file 1). Hospital admissions were analysed in similar groups based on ICD-10 coding (Additional file 1).

Definitions of heat

We defined heat-waves using two heat metrics: daily apparent maximum air temperature (TAppMax) [29] and daily maximum temperature (Tmax). Apparent temperature is a measure of relative discomfort due to combined heat and high humidity. This measure was developed by Steadman [30,31] on the basis of physiologic studies on evaporative skin cooling and can be calculated as a combination of air temperature and dew point. The occurrence of negative dew point values in Australian data makes the use of the commonly used Kalkstein and Valimont [32] apparent temperature formula invalid as it involves taking the square of dewpoint, which introduces a sign error on days of low humidity. In this study we used the Steadman [33] formula. Daily maximum temperature is the highest temperature reached in a given location over a given 24 hour period from midnight to midnight.

We defined single day severe heat events as days where the temperature metric exceeds the 95thpercentile

of the monthly distribution for the study period in that zone. Similarly, three day severe heat events are defined as days when the three day moving average (lags 0–2) exceeds the 95thpercentile. We then defined single day extreme events as days when the temperature metric exceeds the 99thpercentile of the monthly distribution for the study period in that zone, and three day extreme events (or heat-waves), as days when the three day moving average (lags 0–2) exceeds the 99thpercentile. The association between mortality and hospital admissions with severe events provides information on the health effects associated with relatively regular heat events (on average nine events during spring and summer each year), while the association with extreme events provides information on the health effects that occur during less frequent heat events (on average two events per year).

Additional sensitivity analyses which assessed four alternative heat event definitions for associations with mortality from 1) all causes, 2) all cardiovascular disease and 3) all respiratory disease were conducted. These alternative heat-wave definitions were proposed by the Australian Bureau of Meteorology and the NSW Ministry of Health to capture the integrated experience of high temperatures over multiple days. The Excess Heat Factor (EHF) index proposed by Nairn et al. [34] calculates the difference between the 3 day moving average (lags 0–2) of daily average temperature and the prior 30 day moving average (lags 3–32). This is then amplified by the difference of the three day average and the 95th percentile of daily average temperatures. Days with an EHF > 1 are defined as "heat-wave" days (EHF severe). As EHF > 1 occurs relatively frequently we also defined "extreme heat-wave" days as EHF > 80 percentile value (EHF extreme). The third definition which is NSW specific (NSW1) [35] identifies a heat wave as two consecutive days greater than the 95th percentile of maximum temperature for a given month. The fourth definition, also NSW specific (NSW2), proposes that a heat wave is any three day period with greater than the 90th percentile of maximum temperatures for that month in addition to a moving average of greater than 27°C maximum temperature for the 3 day period.

Statistical analyses

We used a time-stratified case-crossover design, in which each case is their own control [36]. Comparisons were made between the 'event' day (the day the case was admitted to hospital or died) and several referent 'non-event' days, with measured and unmeasured potential confounding factors such as age and smoking status controlled by design. The referent days were selected from the same month and year and matched by day of week to the health outcome. This time-stratified method of selecting comparison days ensures unbiased conditional logistic regression estimates and avoids bias resulting from time trends in examination of the environmental exposures [37]. We used conditional logistic regression models to calculate odds ratios (OR) and 95% confidence intervals (CI) for mortality and hospital admission on *extreme* heat days (lag0, lag1, lag2, lag3, lag 0–2 days) compared to other days, adjusted for potential confounders and stratified by zone (ie: including an indicator for zone). The covariates were chosen a priori based on our literature review and were included in all models.

To control for influenza in the model we included a dummy variable for daily counts of influenza admissions (ICD 10: J09 – J18) greater than the 90thpercentile of the Sydney GMR distribution. Single days within an influenza event lasting several days/weeks may dip below the 90thpercentile and we included these days as influenza event days. In addition, an indicator variable was added to the model to account for public holidays. School holidays were also included as an indicator variable when hospital admissions for asthma was the outcome as this group includes a substantial proportion of children. Table 1 summarises the covariates included in the statistical model (Table 1).

Air pollution concentrations may also be elevated during extreme and severe heat events and so potential confounding by air pollution was assessed by conducting sensitivity analyses that included air pollutants (PM_{10} (24 hour average), $PM_{2.5}$ (24 hour average), NO_2 (1 hour maximum) and O_3 (1 hour maximum)) separately in the models. Data on air pollutants were obtained from the NSW Office of Environment and Heritage. Daily region-wide average air pollution concentrations were calculated using daily data from all available monitoring stations in each of the five zones (monitoring stations: Sydney east, n = 4; Sydney west, n = 7; Gosford/Wyong, n = 0; Newcastle, n = 3; Illawarra, n = 3). Gosford/Wyong did not have any monitoring stations so those from the neighbouring city of Newcastle were used.

Table 1 Covariates included in statistical models

Exposure variable	Definition
TempMax (TMax) (°C)	Same day temperature max - lag0, lag1, lag2, lag3, av (lags0-2)
DewPointMax (°C)	Same day dew point temperature max– lag0
Flu Epidemic (0,1)	Binary variable coded as 1 if influenza hospital admission rates were < 90th percentile and otherwise 0
TempAppMax (TAppMax)	Same day Apparent Temperature Max - lag0, lag1, lag2, lag3, av(lags0-3)
School holiday	Binary variable coded as 1 if school holidays and otherwise 0. Only included in childhood asthma models.
Public holiday	Binary variable coded as 1 if public holiday and otherwise 0.
Region	Dummy variable/s for regions

Wilson et al. Environmental Health 2013, 12:98
http://www.ehjournal.net/content/12/1/98

Results

Temperature distributions varied across the five climate zones studied and summary statistics including temperature thresholds are presented in Table 2. The monthly 95thpercentile thresholds in Sydney east ranged from 27.6°C in September to 32.8°C in January, while in Sydney west they ranged from 28.5°C in September to 35.4°C in January. Figure 2 shows heat event days as defined by the 99thpercentile and 95thpercentile for Sydney east for the period 1997 to 2007, as well as the four alternative heat wave definitions.

As the focus of this study is to inform the development of a HHWS, comparisons were made between mortality and hospital admissions associated with rare occurrences of single day and three day extreme heat events, and those of the more frequent single day and three day severe heat events. Results from the statistical model (without controlling for air pollutants) are reported. Including air pollutants in the model had little influence on the mortality results but did influence some results for hospital admissions and these are highlighted in the text.

We found broadly similar associations between mortality and hospital admissions with single day events as well as three day events (lag 0 – 2). The largest single day lag was generally same day (lag0) although this did vary for hospital admissions. The results were also similar for extreme heat events and severe heat events defined using daily maximum temperature and daily maximum apparent temperature. The magnitude of the associations between the rare extreme events (99thpercentile definition) and the more frequent severe events (95thpercentile definition) was also generally similar. The magnitude of the associations was generally similar for all-age and 65 + years. The asthma hospital admissions results are reported for childhood asthma (1–14 years) and adult asthma (15–64 years) rather than all age asthma hospital admissions due to the different epidemiology of asthma in children and adults. The associations for mortality and hospital admissions with severe heat events (95thpercentile threshold) and extreme heat events (99thpercentile threshold) defined using maximum daily temperature are summarized in Table 3 and Table 4 respectively. The results highlighted in the text below are generally for single day severe events as the results for multiple day severe events and single and multiple day extreme events were usually similar. Results are presented in parentheses as percentile threshold, lag: (OR: 95% LCI – 95% UCI).

All-cause mortality had similar magnitude associations with single day and three day extreme and severe events (95thpercentile lag0 (1.06: 1.03 – 1.09)). All cause hospital admissions also showed similar small magnitude associations with three day and single day extreme and severe events (95thpercentile, lag0 (1.02: 1.01 – 1.03)) although only the association with three day extreme events remained when ozone was included in the model (Table 5).

Mortality from disorders of fluid, electrolyte and acid–base balance had large magnitude associations for single day (lag 1 day) and three day extreme and severe events (95thpercentile, lag1 (4.94:2.16 – 11.29)). However, mortality from dehydration was only weakly associated with single day (lag 1 day) and three day severe events (95thpercentile, lag1 (1.54: 0.93 – 2.53)) and not with extreme events. As expected we found large magnitude associations with hospital admissions for heat related injury and illness with single day and three day severe and extreme events (95thpercentile, lag0 (12.05: 8.91 – 16.30)) although controlling for air pollution approximately halved the magnitude of these associations. Three day and single day extreme and severe events showed similar magnitude associations with hospital admissions for dehydration (95thpercentile, lag0 (1.77: 1.60 – 1.97)) and admissions for disorders of fluid, electrolyte and acid base balance (95thpercentile, lag0 (1.27:1.14 – 1.42)), while controlling for ozone marginally reduced the magnitude of these associations.

Cardiovascular mortality was associated with single day and three day extreme and severe events and the magnitude of these associations were larger for multiple day events compared to single day events (95thpercentile, lag0 (1.06: 1.01 – 1.12)) and these results were reflected in ischaemic heart disease mortality (95thpercentile, lag0 (1.07: 1.01 – 1.14)). Cerebrovascular disease mortality was associated with three day extreme events (99thpercentile, lag0-2 (1.15: 0.98 – 1.34)) but not single day extreme events, nor was it associated with three day or single day severe events. We found all cardiovascular admissions were inversely associated with single day extreme and severe heat events lagged 1 day (95thpercentile, lag1(0.96: 0.94 – 0.98) and these inverse associations were reflected in IHD (95thpercentile, lag1(0.96:0.93 – 0.98)) and cerebrovascular disease (95thpercentile, lag1(0.97: 0.88 – 0.97)).

Table 2 Descriptive temperature statistics by region over the study periods (1997 to 2010)

Daily temperature (°C)	Min	Max	Mean	95th percentile range (min – max)	99th percentile range (min – max)
Sydney East	11.3	43.0	21.9	27.6 – 32.8	30.5 – 36.7
Sydney West	10.8	43.3	22.4	28.2 – 35.4	31.5 – 41.4
Newcastle	9.4	42.0	21.6	26.8 – 32.5	29.9 – 36.4
Gosford	11.0	43.3	20.8	26.3 – 29.6	28.6 – 33.9
Illawarra	10.9	41.8	22.1	27.9 – 33.6	31.3 – 37.2

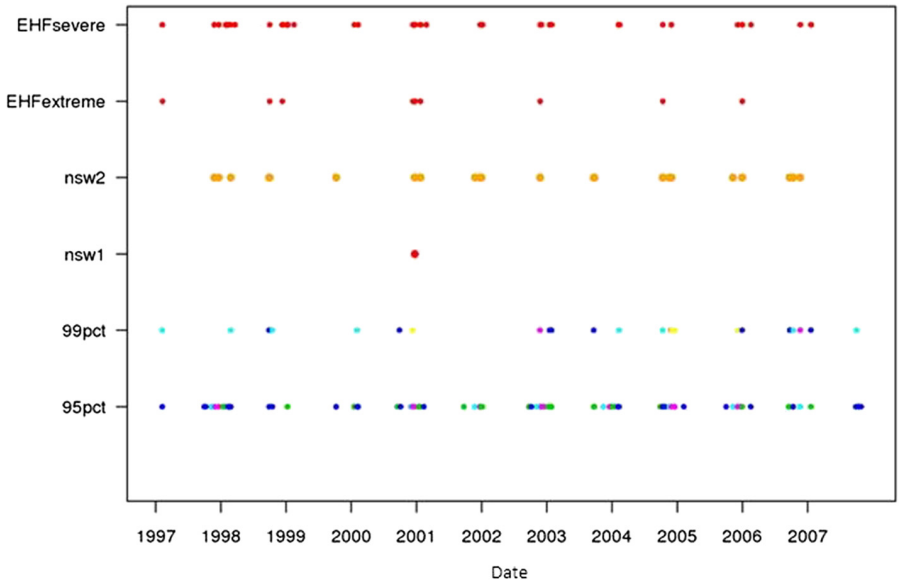


Figure 2 Heat event days as defined by the 99th percentile and 95th percentile for Sydney East for the period 1997 to 2007, as well as the four alternative heat wave definitions.

Table 3 Daily mortality statistics by cause of death in greater metropolitan Sydney regions 28th Feb 1997 – 2007

Cause of death (ICD-10)	Greater Metropolitan Sydney Region		
	Daily deaths by cause		
	Min (Range)	Max (Range)	Median (Range)
All causes (ICD-10 A00 – R99)	0 – 16	14 – 62	4 – 33
Cardiac arrest (ICD-10 I46)	0 – 0	1 – 2	0 – 0
Ischaemic heart diseases (ICD-10 I20 – 125.9)	0 – 0	5 – 16	1 – 6
CVD (ICD-10 I00 – 199.9)	0 – 3	8 – 34	0 – 3
Cardiovascular disease (ICD-10 I01 – I51.9)	0 – 1	6 – 24	1 – 9
Atrial fibrillation (ICD-10 I48 – I48.9)	0 – 0	2 – 3	0 – 0
Myocardial infarct (ICD-10 I21 – I22.9)	0 – 0	5 – 11	0 – 3
Congestive heart failure (ICD-10 I50 – I50.9)	0 – 0	2 – 5	0 – 0
Cerebrovascular disease (ICD-10 I60 – I69.9)	0 – 0	4 – 15	0 – 4
Stroke (ICD-10 I64 – I64.9)	0 – 0	3 – 6	0 – 1
Respiratory disorders (ICD-10 J00 – J99.9)	0 – 0	3 – 6	0 – 3
COPD (ICD-10 J40 – J44.9 or ICD-10 J47 – J47.9)	0 – 0	3 – 6	0 – 1
Pneumonia (ICD-10 J12 – J18.9)	0 – 0	2 – 6	0 – 0
Mental disorders (ICD-F00 – F99.9)	0 – 0	2 – 8	0 – 1
Mental (substance abuse) (ICD-F10 – F19.9)	0 – 0	2 – 4	0 – 0
Dementia (ICD-10 F00 – F03.9)	0 – 0	2 – 8	0 – 0
Diabetes mellitus (ICD-E10 – E14.9)	0 – 0	2 – 4	0 – 0
Schizophrenia (ICD-F20 – F29)	0 – 0	2 – 4	0 – 0
Suicide (ICD-10 ×60 – ×84.9 or Y870)	0 – 0	2 – 4	0 – 0
Dehydration (ICD-10 E86 – E86.9)	0 – 0	1 – 2	0 – 0
Disorders of fluid balance (ICD-10 E86 – E86.9)	0 – 0	1 – 2	0 – 0
Renal failure (ICD-10 N17 – N19.9)	0 – 0	2 – 4	0 – 0

Wilson et al. Environmental Health 2013, 12:98
http://www.ehjournal.net/content/12/1/98

Table 4 Daily hospital admission statistics by condition in greater metropolitan Sydney region

Condition (ICD-10)	Greater Metropolitan Sydney Region		
	Daily admissions by cause		
	Min (Range)	Max (Range)	Median (Range)
All causes (ICD-10 A00 – R99)	24 – 206	113 – 515	63 – 333
Cardiac arrest (ICD-10 I46)	0 – 0	2 – 4	0 – 0
Ischaemic heart diseases (ICD-10 I20 – 125.9)	0 – 4	15 – 37	4 – 17
CVD (ICD-10 I00 – 199.9)	1 – 20	27 – 90	10 – 52
Cardiovascular disease (ICD-10 I01 – I51.9)	0 – 13	23 – 68	7 – 38
Cerebrovascular disease (ICD-10 I60 – I69.9)	0 – 0	7 – 21	1 – 8
Respiratory disorders (ICD-10 J00 – J99.9)	0 – 14	22 – 93	7 – 43
COPD (ICD-10 J40 – J44.9 or ICD-10 J47 – J47.9)	0 – 1	11 – 35	2 – 9
Pneumonia (ICD-10 J12 – J18.9)	0 – 2	8 – 31	1 – 10
Asthma All Ages	0 – 2	8 – 48	1 – 8.5
Mental disorders (ICD-F00 – F99.9)	0 – 9	15 – 56	4 – 25
Mental (substance abuse) (ICD-F10 – F19.9)	0 – 0	6 – 26	1 – 7
Dementia (ICD-10 F00 – F03.9)	0 – 0	3 – 7	0 – 1
Schizophrenia (ICD-F20 – F29)	0 – 0	6 – 20	0 – 6
Renal failure (ICD-10 N17 – N19.9)	0 – 0	4 – 8	0 – 2
Urolithiasis (ICD-10 N20– N23)	0 – 0	6 – 17	0 – 4.5
Dehydration (ICD-10 E86 – E86.9)	0 – 0	4 – 7	0 – 1
Disorders of fluid balance (ICD-10 E87 – E87.8)	0 – 0	4 – 7	0 – 1
Diabetes mellitus (ICD-E10 – E14.9)	0 – 0	6 – 17	1 – 5
Heat related injuries (ICD-10 T67 – T67.9)	0 – 0	2 – 8	0 – 0
Neurotic /Somatic/Stress (ICD-10 F40 – F48)	0 – 0	5 – 23	1 – 3

Table 5 Odds ratios of mortality associated with severe and extreme heat (95th/99th percentile) in Sydney GMR

Deaths		Single day (Lag0) OR (95%CI)		3 day event (Lag0-2) OR (95%CI)	
		95th percentile	99th percentile	95th percentile	99th percentile
All cause	All Age	*1.06 (1.03 - 1.09)*	*1.07 (1.02 – 1.13)*	*1.06 (1.04 - 1.09)*	*1.09 (1.03 – 1.15)*
	65+	*1.05 (1.02 - 1.08)*	1.05 (0.99 - 1.12)	*1.05 (1.02 - 1.08)*	*1.09 (1.03 - 1.15)*
All cardiovascular	All Age	*1.06 (1.01 - 1.12)*	*1.07 (0.97 - 1.18)*	*1.11 (1.06 - 1.17)*	*1.17 (1.06 - 1.29)*
	65+	*1.06 (1.01 - 1.12)*	1.09 (0.98 - 1.20)	*1.11 (1.05 - 1.17)*	*1.20 (1.08 - 1.32)*
IHD	All Age	*1.07 (1.01 - 1.14)*	*1.12 (1.01 - 1.26)*	*1.12 (1.06 - 1.18)*	*1.15 (1.03 - 1.29)*
	65+	*1.06 (1.00 - 1.13)*	*1.13 (1.01 - 1.27)*	*1.10 (1.04 - 1.18)*	*1.17 (1.04 - 1.32)*
All respiratory	All Age	*1.14 (1.04 - 1.24)*	1.08 (0.91 - 1.29)	1.08 (0.99 - 1.18)	0.99 (0.83 - 1.19)
	65+	*1.13 (1.03 - 1.23)*	1.09 (0.91 - 1.30)	1.07 (0.97 - 1.17)	0.97 (0.80 - 1.18)
Diabetes	All Age	*1.22 (1.03 - 1.46)*	1.13 (0.81 - 1.58)	1.16 (0.97 - 1.39)	0.88 (0.59 - 1.32)
	65+	*1.26 (1.05 - 1.51)*	1.20 (0.83 - 1.70)	1.14 (0.94 - 1.39)	0.92 (0.60 - 1.40)
Renal failure	All Age	1.18 (0.95 - 1.45)	1.14 (0.75 - 1.73)	1.09 (0.88 - 1.35)	0.89 (0.54 - 1.46)
	65+	1.17 (0.94 - 1.45)	1.08 (0.69 - 1.67)	1.05 (0.83 - 1.31)	0.90 (0.54 - 1.50)
Mental disorders	All Age	1.10 (0.95 - 1.27)	0.88 (0.64 - 1.20)	1.04 (0.90 - 1.21)	0.93 (0.68 - 1.28)
	65+	1.07 (0.91 - 1.27)	0.82 (0.58 - 1.16)	0.98 (0.83 - 1.16)	0.81 (0.56 - 1.66)
Disorders of fluid balance	All Age	1.10 (0.46 - 2.63)	1.98 (0.49 - 8.06)	*2.53 (1.18 - 5.41)*	*5.63 (1.62 - 19.56)*
	65+	1.02 (0.39 - 2.60)	2.37 (0.55 - 10.17)	*2.37 (1.00 - 5.61)*	0.66 (0.23 - 1.97)

Respiratory mortality was associated with single day and three-day severe events, but not with extreme events at (95^{th}percentile, lag0 (1.14: 1.04 – 1.24). Pneumonia mortality was associated with single day extreme and severe events lag2 days (95^{th}percentile, lag2 (1.25:1.05 – 1.49) but not three day severe or extreme events. COPD mortality was associated with single day severe events (95^{th}percentile, lag0 (1.18:1.05 – 1.33), with a weaker association with three day severe events, while extreme events were not associated with COPD. Hospital admissions for all respiratory disorders were not associated with extreme events although there was a small magnitude association with three day and single day severe events (95^{th}percentile, lag 0 (1.03:1.01 – 1.05)). While these results were reflected in pneumonia admissions and COPD admissions the associations were generally weak and inconsistent. These weak associations with respiratory outcomes were generally eliminated when air pollutants were included in the models, particularly ozone and nitrogen dioxide. We found in inverse association between childhood asthma admissions and single day severe and extreme events (95^{th}percentile, lag0 (0.94: 0.89 – 1.01)). Controlling for air pollutants generally magnified this inverse association, particularly for ozone and nitrogen dioxide, such that three day severe and extreme events were also inversely associated with childhood asthma (heat + ozone model; 95^{th}percentile, lag0-2 (0.91: 0.84 – 0.98)). Adult asthma admissions were inversely associated with both single day and three day extreme and severe events and controlling for air pollution generally give similar results (95^{th}percentile, lag0 (0.92: 0.84 – 1.02)).

Mortality from renal failure was not associated with extreme or *severe* events. Three day and single day extreme and severe events were associated with hospital admissions for disorders of the renal system with the strongest single day associations at lag1 and lag2 days (95^{th}percentile, lag1 (1.08: 1.05 – 1.12)). Controlling for air pollutants in the model reduced the magnitude and strength of these associations. Hospital admissions for urolithiasis (kidney stones) was associated with three day and single day extreme and severe events (95^{th}percentile, lag1 (1.18:1.02 – 1.25)). Admissions for renal failure were associated with three day extreme and severe events as well as single day severe events but not extreme events (95^{th}percentile, lag0 (1.26: 1.16 – 1.38)) although controlling for air pollutants reduced the magnitude of the associations with severe events and eliminated the association with three day extreme events.

Diabetes mortality was associated with single day severe events (95^{th}percentile, lag0 (1.22:1.03 – 1.46) and weakly associated with three day severe events, but was not associated with extreme events. Hospital admissions for diabetes were associated with similar magnitude associations with three day and single day extreme and severe events (95^{th}percentile, lag0 (1.12: 1.06 – 1.18)). Controlling for air pollutants generally gave similar magnitude associations for diabetes hospitalisations although the strength of these associations was reduced with three day extreme and severe events.

Deaths from all mental disorders were not associated with extreme or severe events for the all-ages group but were associated with single day extreme and severe events in the under 65 years age group lag 2 days (99^{th}percentile, lag2 (2.84: 1.24 – 6.52) and 95^{th}percentile, lag2 (1.53: 1.10 – 2.14)). Deaths from mental disorders (substance abuse) showed similar magnitude associations with single day and three day severe events (95^{th}percentile, lag0 (1.38: 1.01 – 1.88)) as well as single day extreme events at lag1 and lag 2 days (99^{th}percentile, lag1 (2.30: 1.19 – 4.44)) but not three day extreme events. Deaths from suicide were not associated with extreme events although there was a weak association with single day severe events lagged 3 days which only occurred in the under 65 year age group (95^{th}percentile, lag3 (1.19: 0.98 – 1.44)). Mortality from dementia was inversely associated with single day severe and extreme events lagged 3 days, but not three day events (95^{th}percentile, lag3, (0.83: 0.68 – 0.99)). Similarly, organic mental disorders were inversely associated with single day severe and extreme events lagged 3 days, but not three day events (95^{th}percentile, lag3 (0.85: 0.71 – 1.02)).

Hospital admissions for all mental and behavioural disorders showed small magnitude associations with three day and single day extreme and severe events with the strongest single day association at lag1 days (95^{th}percentile, lag1 (1.02:1.00 – 1.05) (Table 6). While these results were reflected in various categories of hospital admissions for mental and behavioural disorders the associations were generally weak and inconsistent. Hospital admissions for mood disorders were weakly associated with three day and single day severe events (95^{th}percentile, lag1 (1.06:1.00 – 1.12)). Admissions for mental disorders related to substance abuse were weakly associated with three day and single day extreme events (99^{th}percentile, lag0 (1.08: 0.97 – 1.20)). Admissions for dementia were associated with single day severe events (95^{th}percentile, lag0 (1.14:0.99 – 1.31)) as were admissions for neurotic/ somatic/ stress (95^{th}percentile, lag0, (1.06:0.99 – 1.14)). There was no consistent pattern of change in results for mental disorder admissions when air pollutants were included in the model with the magnitude of some associations decreasing, for example, dementia with ozone included in model (95^{th}percentile, lag0 (1.08:0.92 – 1.28)). While the magnitude of some mental disorder admissions associations increased, for example, mental disorders related to substance abuse and PM2.5 (heat model: 99^{th}percentile, lag0-2 (1.06: 0.96 – 1.18); heat + PM2.5 model: 99^{th}percentile, lag0-2 (1.11: 0.99 – 1.25)). As hospital admissions for mental and behavioural disorders are unlikely to be causally related to air pollution,

Wilson et al. Environmental Health 2013, 12:98
http://www.ehjournal.net/content/12/1/98

Table 6 Odds ratios of hospitalisations associated with severe and extreme heat (95th/99th percentile) in Sydney GMR

Hospital admissions		Single day (Lag0) OR (95%CI)		3 day event (Lag0-2) (95%CI)	
		95th percentile	99th percentile	95th percentile	99th percentile
All cause	All Age	*1.02 (1.01 - 1.03)*	*1.01 (1.00 - 1.02)*	*1.02 (1.01 - 1.03)*	*1.03 (1.02 - 1.05)*
	65+	*1.04 (1.03 - 1.06)*	*1.03 (1.01 - 1.06)*	*1.03 (1.02 - 1.04)*	*1.05 (1.03 - 1.07)*
All cardiovascular	All Age	1.01 (0.99 - 1.03)	*0.98 (0.94 - 1.03)*	0.98 (0.96 - 1.00)	0.98 (0.94 - 1.03)
	65+	1.01 (0.93 - 0.98)	0.98 (0.93 - 1.04)	0.97 (0.95 - 1.00)	0.96 (0.91 - 1.02)
IHD	All Age	1.04 (0.86 - 1.24)	*1.00 (0.93 - 1.06)*	0.99 (0.96 - 1.00)	1.00 (0.93 - 1.06)
	65+	*1.02 (0.98 - 1.06)*	1.02 (0.94 - 1.11)	*0.99 (0.95 - 1.03)*	0.99 (0.91 - 1.07)
All respiratory	All Age	1.03 (0.96 - 1.06)	0.98 (0.94 - 1.03)	*1.03 (1.01 - 1.05)*	1.01 (0.97 - 1.06)
	65+	*1.07 (1.04 - 1.10)*	1.03 (0.96 - 1.09)	*1.08 (1.04 - 1.11)*	*1.07 (1.01 - 1.14)*
Pneumonia	All Age	*1.04 (1.00 - 1.08)*	0.97 (0.89 - 1.06)	*1.05 (1.01 - 1.09)*	1.10 (1.01 - 1.20)
	65+	*1.06 (1.00 - 1.12)*	1.01 (0.90 - 1.13)	*1.09 (1.03 - 1.15)*	*1.12 (1.00 - 1.25)*
Asthma	1 - 14 yrs	*0.94 (0.89 - 1.01)*	*0.92 (0.80 - 1.04)*	*0.93 (0.87 - 0.98)*	0.89 (0.79 - 1.01)
	15 - 64 yrs	*0.96 (0.88 - 1.04)*	*0.86 (0.72 - 1.04)*	0.95 (0.87 - 1.04)	0.80 (0.66 - 0.97)
COPD	All Age	*1.06 (1.01 - 1.10)*	1.05 (0.97 - 1.14)	*1.06 (1.02 - 1.11)*	1.06 (0.97 - 1.15)
	65+	*1.07 (1.03 - 1.13)*	1.03 (0.94 - 1.14)	*1.07 (1.02 - 1.13)*	1.03 (0.93 - 1.14)
Diabetes	All Age	*1.12 (1.06 - 1.18)*	*1.16 (1.03 - 1.30)*	*1.07 (1.01 - 1.04)*	*1.14 (1.01 - 1.29)*
	65+	*1.12 (1.03 - 1.21)*	1.18 (0.99 - 1.39)	*1.06 (0.97 - 1.15)*	1.15 (0.95 - 1.37)
All renal disorders	All Age	*1.04 (1.01 - 1.10)*	*1.01 (0.94 - 1.08)*	*1.14 (1.04 - 1.11)*	*1.19 (1.12 - 1.27)*
	65+	*1.09 (1.04 - 1.14)*	1.01 (0.91 - 1.12)	*1.12 (1.07 - 1.17)*	*1.26 (1.15 - 1.40)*
Urolithiasis	All Age	1.01 (0.95 - 1.07)	1.05 (0.97 - 1.14)	*1.15 (1.08 - 1.22)*	*1.32 (1.18 - 1.50)*
	65+	1.07 (0.91 - 1.25)	0.90 (0.63 - 1.29)	*1.31 (1.12 - 1.53)*	*1.79 (1.34 - 2.40)*
Mental disorders	All Age	1.01 (0.98 - 1.04)	*1.08 (0.97 - 1.20)*	*1.03 (1.00 - 1.05)*	1.05 (0.99 - 1.10)
	65+	1.03 (0.96 - 1.09)	0.94 (0.81 - 1.09)	*1.09 (1.02 - 1.16)*	1.05 (0.91 - 1.20)
Dehydration	All Age	*1.78 (1.60 - 1.97)*	*2.16 (1.80 - 1.60)*	*1.72 (1.56 - 1.91)*	*2.32 (1.91 - 2.81)*
	65+	*1.85 (1.63 - 2.09)*	*2.08 (1.64 - 2.64)*	*1.78 (1.51 - 2.02)*	*2.15 (1.68 - 2.76)*
Heat related illness and injury	All Age	*12.05 (8.91 - 16.30)*	*18.76 (11.38 - 30.93)*	*1.77 (1.60 - 1.97)*	*10.35 (6.09 - 17.59)*
	65+	*19.47 (11.71 - 32.36)*	No data	*7.05 (4.69 - 10.60)*	No data
Disorders of fluid balance	All Age	*1.27 (1.14 - 1.42)*	*1.50 (1.20 - 1.86)*	*1.19 (1.06 - 1.33)*	*1.49 (1.19 - 1.86)*
	65+	*1.25 (1.10 - 1.42)*	*1.51 (1.18 - 1.93)*	*1.16 (1.02 - 1.32)*	*1.50 (1.16 - 1.94)*

the change in the association with these outcomes suggests co-linearity may be influencing the ability of the model to assess the independent effects of heat and air pollution.

Additional analyses that assessed the four alternative heatwave definitions designed to integrate the effect of high temperature over multiple days with all-cause, cardiovascular and respiratory mortality gave similar magnitude associations with these outcomes as the simpler heat wave definitions reported here.

Discussion

We found that single day and three day severe heat events were associated with increased mortality and hospital admission in the Sydney GMR as well as less frequent and more intense extreme heat events. Our study found broadly similar results irrespective of the definition and metric (maximum temperature or apparent temperature) used to define heat events, consistent with the finding of Khalaj et al. and Barnett et al. [9,38]. While air pollution does not appear to confound the association between mortality and heat events, we found evidence of confounding by air pollution with some hospital admissions outcomes, particularly respiratory outcomes and ozone. The associations for some hospital admissions outcomes that were influenced by controlling for air pollutants in the statistical model are unlikely to be causally related to air pollution, for example heat related injury. This suggests that co-linearity between air pollutants and heat events may be influencing the ability of the model to assess the independent effects of heat and air pollution.

Our findings suggest mortality from disorders associated with fluid imbalance, as well as cardiovascular disease, pneumonia, COPD, diabetes, and mental disorders related to substance abuse are associated with heat events. As expected heat events were associated with substantial increases in hospital admissions for dehydration, other disorders of fluid, electrolyte and acid base balance, and heat related injury. We also found that heat events were associated with hospital admissions for renal disease, diabetes, and mental and behavioural disorders. These increases in mortality and hospitalisation during heat events place an additional load on already stretched emergency services, hospitals and the health system.

As expected hospital admission for heat related injury, disorders of fluid balance and dehydration had large magnitude associations with both single day and three day extreme and severe events. While deaths from dehydration were only weakly associated with severe events we found a large magnitude association with other fluid disorders and both severe and extreme events.

We found mortality and hospital admissions for mental disorders were associated with both single day and three day extreme and severe heat events. Deaths from all mental and behavioural disorders were associated with single day extreme and severe events in the under 65 year age group but not the all-age group. This was reflected in associations for hospital admissions for all mental and behavioural disorders with extreme and severe events. There was some consistency in the results for mortality and hospital admissions for mental disorders related to substance abuse with both outcomes associated with extreme events and mortality associated with severe events. The results for other sub categories of mortality and admissions for mental disorders were mixed. While we found inverse associations for mortality from organic mental disorders and dementia mainly with single day severe events these outcomes had very low daily counts with limited statistical power. We found weak positive associations with hospital admissions for mood, dementia, and somatic/neurotic/stress disorders and mainly single day severe events. Associations between heat events and mental and behavioural disorders have been found elsewhere [11,39]. A study in Adelaide, Australia found heat-waves were associated with mortality due to a range of mental health disorders [40]. The increased intensity and duration of heat-waves in Adelaide may explain the weaker associations with mortality we found in Sydney. Psychotropic drugs are commonly used by people with mental and behavioural disorders and these drugs may interfere with body temperature regulation to exacerbate the effects of heat-waves [39]. In Khalaj's recent Sydney study, an association between hospital admission for mental and behavioural disorders, as both a primary diagnoses (compared with persons admitted for other conditions) and as an underlying diagnosis, (compared with persons admitted who did not have the condition) [9] was found. We found a weak association between suicide in the under 65 year age group and single day severe events. This result should be treated with caution as many papers have explored links between suicide and temperature and produced conflicting results [41-44]. Some studies have found a decreased suicide risk with increasing temperature [41], while in others an increased risk was observed [42]. One study found a U-shaped response with elevated suicide risk on extremely cold and warm days [43]. Our study contributes to the emerging literature in this area.

We found hospital admissions for renal failure, diabetes and urolithiasis were associated with extreme and severe heat events. While mortality from renal failure was not associated with heat events the small number of deaths coded to renal failure limit the statistical power of the analysis to find an association if one was present. A recent study in France found heat waves were associated with renal failure and that these effects were larger in elderly patients [45]. It is reasonable to assume that these associations may be related to increased fluid loss and subsequent altered fluid balance and dehydration.

While diabetes hospital admissions were associated with singe day and three day severe and extreme events, diabetes mortality was only weakly associated with single day and three day severe events. Once again the small number of deaths coded to diabetes limits the statistical power of the analysis to find an association if one were present. Disturbances to the fluid balance resulting from high temperature could also be a contributing factor to the observed associations with diabetes [46].

Cardiovascular disease mortality and the related outcomes of IHD and cerebrovascular disease were associated with single day and three severe and extreme events. Conversely we found weak inverse associations with single day extreme and severe events for all cardiovascular admissions, as well as IHD and cerebrovascular disease admissions. This suggests that heat events may lead to a rapid deterioration in existing cardiovascular disease resulting in death, and that this may occur during rare extreme events as well as the more frequent severe events. European studies have similarly found an association between heat-waves and cardiovascular mortality but not for hospital admissions where cardiovascular disease was the primary cause [47-51]. Khalaj's 2010 Sydney case-only study did not find any association between heat-waves and hospital admissions for cardiovascular disease as the primary diagnosis (compared with persons admitted for other conditions) but did find an association with cardiovascular disease as an underlying diagnosis (compared with persons admitted who did not have the condition). This suggests that an underlying condition of cardiovascular disease

Wilson et al. Environmental Health 2013, 12:98
http://www.ehjournal.net/content/12/1/98

modifies the effect of heat on hospital admissions for a range of other conditions [9,52].

Whilst respiratory mortality was associated with single day and three day severe events, it was generally not associated with extreme events, although once again the small number of deaths coded to respiratory mortality limits the statistical power of the analysis to find an association if one were present, especially for less frequent *extreme* events. While respiratory hospital admissions were associated with heat events all associations were eliminated when we controlled for the effects of air pollution, particularly ozone. Air pollution is associated with respiratory morbidity [53-55] and studies suggest that ozone can confound the association with heat for cardiovascular and respiratory outcomes [56]. Air pollution concentrations are influenced by weather and ozone in particular is correlated with temperature [57]. Heat and air pollution may have independent effects of respiratory outcomes however colinearity may be influencing the ability of our statistical model to assess the independent effects of heat and air pollution and work is starting to emerge exploring the issue of confounding and effect modification of heat by air pollution [58]. More broadly, our finding that the effects of heat on some hospital admissions disease groups are sensitive to the effect of air pollution, particularly O_3 and NO2, is consistent with previous work that identified this phenomenon in the 2003 heatwave in France [59].

We found an inverse association with single day and three day extreme events for childhood asthma, and for both severe and extreme events for adult asthma. Asthma time series data, particularly for children, is difficult to model [60]. Probably because children are more likely to experience frequent respiratory epidemics and because school holidays influence hospital admission rates [61]. The protective effect for asthma may also be due to the warm moist air associated with heat events in Sydney's temperate climate. Asthma is known to by triggered by inhaling cold dry air, and in an experimental study, warm moist air has been shown to protect the airway disruption associated with exercise [62]. It is also possible that on very hot days, people with asthma stay inside more, and are reluctant to exercise due to concerns for triggering an asthma attack. A recent Sydney case-only study found heat waves were associated with increased hospital admissions with respiratory disease as an underlying diagnosis (compared with persons admitted who did not have the condition) suggesting that an underlying condition of respiratory disease modifies the effect of heat on hospital admissions for a range of other conditions [9].

Our sensitivity analyses suggest that, at least for mortality, the use of heatwave metrics designed to reflect multiple day events that meet detailed heatwave criteria produce broadly similar results to those found with single day heat metrics such as maximum daily air temperature, or a simple 3 day moving average.

The key strengths of this study include our use of long mortality and hospital admission time series, the wide range of diagnosis groups assessed our adjustment for the potential confounding effects of humidity and our assessment of the influence of air pollution on the heat associations. As with similar ecological time series studies we were limited by the lack of more detailed individual level data on exposure including other potential confounding factors such as the use of air conditioning during heat events, Similar to many other studies of the effects of heat on mortality or hospitalisation we used a time stratified case crossover analysis that has the benefit of controlling for the effects of unmeasured confounding factors that do not vary over time by design. An alternative statistical approach would be to use Poisson regression time series analysis. A recent paper by Guo et al. suggests that case crossover designs may not control for autocorrelation as effectively as time series analysis when assessing air pollution effects on mortality due in part to lack of control for seasonal effects. However, the generalizability of these concerns to our study assessing the effects of extreme heat in interrupted time series analyses of spring and summer months, and to our study location is unclear [63]. The different exposure metrics, different lag effects and large range of mortality and hospitalisation outcomes assessed in our study introduces the potential problem of multiple comparisons. This results in an increased probability of detecting a statistically significant finding by chance alone and the greater potential for spurious results to be reported. One approach to multiple testing suggests using more stringent P-values to account for the larger number of comparisons being conducted, although this can cause more problems than it intends to solve and no correction for multiple comparisons was made in our study [64]. Although this increased the chances of finding spurious associations, it reduced the chances of missing any important associations. Our results summary focuses on consistent lag effects rather than statistically significant effects with erratic lag structures.

This study highlights the increased risk of hospitalisation and death during heat events from cardiovascular disease, diabetes and mental health disorders. These chronic conditions are an important issue in today's society [65]. The aged population are at increased risk of heat-related illness from a range of factors including decreased thermoregulation and an increased prevalence of chronic diseases disease, and these factors contribute to the increased risk of heat related death [66,67]. Many health agencies including the NSW Ministry of Health have implemented programs to improve management of chronic disease in the hope of reducing potentially avoidable admissions [68]. This study provides similar results to research conducted

in the Northern Hemisphere regarding the effect of heatwaves on levels of mortality and morbidity and outlines the importance of including HHWS and heat management plans as part of chronic disease management.

Conclusions

The purpose of this study was to provide data in the Australian context which will inform the development of a Heatwave Health Warning System for NSW. The findings of this study indicate that days with unusually high ambient temperatures increase mortality and hospitalisation from a range of conditions. Ebi and colleagues highlighted the costs and benefits of the establishment of a HHWS in Philadephia, with the benefits of implementing a HHWS far outweighing the costs [6]. We identified increases in mortality and morbidity for a range of conditions on both single and three day severe and extreme events, and a HHWS in NSW based on these thresholds would need to be activated on average one or two times per year for extreme events and around ten times per year for more frequent severe events. Compliance with a HHWS is a critical factor in maximising the protective effect of the warning system and frequent alerts may reduce compliance in the target population. It is recommended that in NSW the HHWS be implemented at the first one or two occurrences of temperatures above the 95th percentile of average monthly temperature (severe event level) from the beginning of spring (September) and thereafter at all occurrences of temperatures above the 99th percentile of average monthly temperature (extreme event level) until the end of summer (February) in order to maximise health protection and reduce complacency. Public health messages associated with all heat health warnings should emphasise the importance of maintaining hydration, particularly for people who have chronic diseases whether the temperature be severe or extreme.

Additional file

Additional file 1: ICD-10 Codes used in analysis.

Abbreviations

BoM: Bureau of meteorology; COPD: Chronic obstructive pulmonary disease; GMR: Greater metropolitan region; HHWS: Heatwave health warning system; ICD – 10: International classification of diseases (Version 10); LGA: Local government area; NSW: New South Wales, Australia; SLA: Statistical local area; TMax: Maximum daily temperature; T(App)Max: Apparent maximum daily temperature.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

GM, FJ and BJ conceived and developed the overall study and conducted the data reviews. LW helped develop the study, assisted with data analysis and prepared the manuscript. IH developed the analysis methods and conducted data analysis. HA-Y conducted data analysis. RB and CG critically reviewed the manuscript and provided advice during study development. All authors read and approved the final manuscript.

Acknowledgments

The authors would like to acknowledge the contributions made by Glenis Lloyd, NSW Health and the members of the Heatwave Health Warning System steering Committee. Funding for this project was provided through NSW Health and Department of Environment Climate Change and Water.

Author details

[1]Faculty of Health Science, University of Sydney, Sydney, Australia. [2]School of Science and Health, University of Western Sydney, Sydney, Australia. [3]University Centre for Rural Health – North Coast, University of Sydney, Sydney, Australia. [4]North Coast Public Health Unit, Mid North Coast Local Health District, New South Wales, Australia. [5]National Centre for Epidemiology and Population Health, Australian National University, Acton, Australia. [6]Menzies Research Institute, University of Tasmania, Hobart, Australia. [7]Centre for Epidemiology and Research, NSW Health, Sydney, Australia. [8]Health Protection New South Wales, Sydney, Australia. [9]NSW Cancer Institute, Sydney, Australia. [10]School of Public Health and Community Medicine, University of New South Wales, Sydney, Australia. [11]Centre for Research, Evidence Management and Surveillance, South Western Sydney Local Health District, Sydney, Australia.

Received: 8 January 2013 Accepted: 23 October 2013
Published: 15 November 2013

References

1. Intergovernmental Panel on Climate Change (IPCC): **Summary for Policymakers**. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Edited by Solomon SD, Qin M, Manning Z, Chen M, Marquis KB, Avery M, Tignor, Miller HL. 2007:5. ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-spm.pdf.
2. Kovats RS, Hajat S: **Heat stress and public health: a critical review.** *Annu Rev Public Health* 2008, **29:**41–55.
3. Kilbourne EM: **Heat-related illness: current status of prevention efforts.** *Am J Prev Med* 2002, **22:**328–329.
4. Bernard SM, McGeehin MA: **Municipal heat wave response plans.** *Am J Public Health* 2004, **94:**1520–1522.
5. Nogueira PJ: **Examples of heat health warning systems: Lisbon's ICARO surveillance system, summer of 2003**. In *Extreme Weather Events and Public Health Responses.* Edited by Kirch WMB, Bertollini R. Darmstadt, Germany: Springer-Verlag; 2005:141–160.
6. Ebi K, Tiesberg TJ, Kalkstein LS, Robinson L, Weiher RF: **Heat watch/warning systems save lives: estimated costs and benefits for Philadelphia 1995 – 1998.** *Am Meteorol Soc* 2004**:**1067–1073.
7. Sheridan SC, Kalkstein LS: **Progress in heat watch-warning system technology.** *Bull Am Meteorol Soc* 2004, **85:**1931–1941.
8. Anderson BG, Bell ML: **Weather-related mortality: How heat, cold and heat waves affect mortality in the United States.** *Epidemiology* 2009, **20:**205–213.
9. Khalaj B, Lloyd G, Sheppeard V, Dear K: **The health impacts of heatwaves in five regions of New South Wales, Australia: a case-only analysis.** *Int Arch Environ Health* 2010, **83:**833–842.
10. Garssen J, Harmsen C, de Beer J: **The effect of the summer 2003 heat wave on mortality in the Netherlands.** *Eurosurveillance* 2005, **10:**pii-577.
11. Johnson H, Kovats RS, McGregor G, Stedman J, Gibbs M, Walton H: **The impact of the 2003 heat wave on mortality and hospital admissions in England.** *Health Stat Q* 2005, **25:**6–11.
12. Fouillet A, Rey G, Laurent F, Pavillion G, Bellec S, Guihenneuc-Jouyaux C, Clavel J, Jougla E, Hemon D: **Excess mortality related to the august 2003 heat wave in France.** *Int Arch Occup Environ Health* 2006, **80:**16–24.
13. Knowlton K, Rotkin-Ellman M, King G, Margolis HG, Smith D, Solomon G: **The 2006 California heat wave: impacts on hospitalisations and emergency department visits.** *Environ Health Perspect* 2009, **94:**1520–1522.
14. Hayhoe K, Sheridan S, Kalkstein L, Greene S: **Climate change, heat waves, and mortality projections for Chicago.** *J Great Lakes Res* 2010, **36:**65–73.
15. Kalkstein L, Sheridan S: **The social impacts of the heat–health watch/warning system in Phoenix, Arizona: assessing the perceived risk and response of the public.** *Int J Biometeorol* 2007, **52:**43–55. doi:10.1007/s00484-006-0073-4.

16. Guest CS, Willson K, Woodward AJ, *et al*: **Climate and mortality in Australia: retrospective study, 1979–1990, and predicted impacts in five major cities in 2030.** *Clim Res* 1999, **13:**1–15.
17. McMichael AJ: **Human health and climate change in oceania: a risk assessment**. In *Commonwealth Department of Ageing*. Canberra, ACT. Australia: Commonwealth Department of Ageing; 2003.
18. Nitschke M, Tucker GR, Bi P: **Morbidity and mortality during heatwaves in metropolitan Adelaide.** *Med J Aust* 2007, **187:**662–665.
19. Bi P, Parton KA, Wang J, Donald K: **Temperature and direct effects on population health in Brisbane, 1986–1995.** *J Environ Health* 2008, **70:**48–53.
20. Vaneckova P, Beggs PJ, Jacobson CR: **Spatial analysis of heat-related mortality among the elderly between 1993 and 2004 in Sydney, Australia.** *Soc Sci Med* 2010, **70:**293–304.
21. Hansen AL, Bi P, Ryan P, Nitschke M, Pisaniello D, Tucker G: **The effects of heat waves on hospital admissions for renal disease in a temperate city of Australia.** *Int J Epidemiol* 2008, **37:**1359–1366.
22. Australian Bureau of Meteorology: 2012. http://reg.bom.gov.au/announcements/media_releases/nsw/2004febnsw.shtm.
23. Australian Bureau of Statistics: *Causes of Death Australia (3303.0). Australian Bureau of Statistics*. Commonwealth of Australia; 2011.
24. NSW Department of Health: *Inpatient Statistics Collection (ISC)*. Sydney Australia: NSW Government; 2010.
25. Wilmore A: *Population Weighted Concordance Between ASGC Versions Between 1998 and 2005 with ASGC 2006, NSW Health Department, 2006*. Sydney Australia: NSW Government; 2006.
26. National Climate Centre of the Bureau of Meteorology: *National Climate Centre of the Bureau of Meteorology: Daily or three hourly weather data for Bureau of Meteorology stations*. Australia: Government of Victoria; 2010.
27. Dear K, Ranmuthugala G, Kjellström T, Skinner C, Hanigan I: **Effects of temperature and ozone on daily mortality during the August 2003 heat wave in France.** *Arch Env Occup Health* 2005, **60:**205–212.
28. Australian Bureau of Statistics, 2007: *Australian Bureau of Statistics. Australian Census 2006 Basic Community Profiles for Census Collector's Districts Data Pack (Cat No. 2069.0.30.001 Second Release)*. Canberra, Australia: Commonwealth of Australia; 2007.
29. Nicholls N, Skinner C, Loughnan M, Tapper N: *Temperature Thresholds Associated with Increased Mortality in ten Major Population Centres in Rural Victoria*. Monash University: Melbourne; 2009.
30. Steadman RG: **The assessment of sultriness: part I: a temperature-humidity index based on human physiology and clothing science.** *J Appl Meteorol* 1979, **18:**861–873.
31. Steadman RG: **The assessment of sultriness: part II: effect of wind, extra radiation. And barometric pressure on apparent temperature.** *J Appl Meteorol* 1979, **18:**874–884.
32. Kalkstein LS, Valimont KM: **An evaluation of summer discomfort in the United States using a relative climatological index.** *Bulletin of the American Meteorological Society* 1986, **7:**842–848.
33. Steadman RG: **Norms of apparent temperature in Australia.** *Aust Meteorol Mag* 1994, **43**. http://www.bom.gov.au/amm/docs/1994/steadman.pdf.
34. Nairn J, Fawcett R, Ray D: *Defining and Predicting Excessive Heat Events, a National System. Centre for Australian Weather and Climate Research Modelling Workshop 2009*. Melbourne: Modelling & Understanding High Impact Weather; 2009. http://www.cawcr.gov.au/events/modelling_workshops/workshop_2009/papers/NAIRN.pdf.
35. Broome R: *NSW Ministry of Health (pers communication)*. NSW Health; 2011.
36. Jaakkola JJ: **Case-crossover design in air pollution epidemiology.** *Eur Respir J Suppl* 2003, **40:**81s–5s.
37. Janes H, Sheppard L, Lumley T: **Case-crossover analyses of air pollution exposure data: referent selection strategies and their implications for bias.** *Epidemiology* 2005, **16:**717–26.
38. Barnett AG, Tong S, Clements AC: **What measure of temperature is the best predictor of mortality?** *Environ Res* 2010, **110:**604–11.
39. García-Pina R, Tobías Garcés A, Sanz Navarro J, Navarro Sánchez C, García-Fulgueiras A: **Effect of weather temperature on hospital emergencies in the Region of Murcia, Spain, throughout the 2000–2005 and its use in epidemiological surveillance.** *Rev Esp Salud Publica* 2008, **82:**153–66.
40. Hansen A, Bi P, Nitschke M, Ryan P, Pisaniello D, Tucker D: **The effects of heatwaves on mental health in a temperate Australian city.** *Environ Health Perspect* 2008, **116:**1369–1375.
41. Dixon P, Kalkstein A: **Climate-suicide relationships: a research problem in need of geographic methods and cross-disciplinary perspectives.** *Geography Compass* 2009, **3:**1–14.
42. Tsai J-F: **Socioeconomic factors outweigh climate in the regional difference of suicide death rate in Taiwan.** *Psychiatry Res* 2010, **179:**212–6.
43. Kim Y, Kim H, Kim D-S: **Association between daily environmental temperature and suicide mortality in Korea (2001–2005).** *Psychiatry Res* 2011, **186:**390–6.
44. Wang Y, Wang D, Wang X: **Suicide and meteorological factors in Huhhot, Inner Mongolia.** *Crisis: The Journal of Crisis Intervention and Suicide Prevention* 1997, **18:**115–7.
45. Josseran L, Caillère N, Brun-Ney D, Rottner J, Filleul L, Brucker G, Astagneau P: **Syndromic surveillance and heat wave morbidity: a pilot study based on emergency departments in France.** *BMC Med Inform Decis Mak* 2009, **9:**14. doi:10.1186/1472-6947-9-14.
46. Timiras P: *Physiological Basis of Ageing and Geriatrics*. 4th edition. CRC Press; 2007. ISBN ISBN-10: 0849373050.
47. Hausfater P, Doumenc B, Chopin S, Le Manach Y, Santin A, Dautheville S, Patzak A, Hericord P, Mégarbane B, Andronikof M, Terbaoui N, Riou B: **Elevation of cardiac troponin I during non-exertional heat-related illnesses in the context of a heatwave.** *Crit Care* 2010, **14:**R99.
48. Mastrangelo G, Hajat S, Fadda E, Buja A, Fedeli U, Spolaore P: **Contrasting patterns of hospital admissions and mortality during heat waves: are deaths from circulatory disease a real excess or an artifact?** *Med Hypotheses* 2006, **66:**1025–1028.
49. Diaz J, Linares C: **Impact of extreme temperatures on daily mortality in Madrid (Spain) among the 45–64 age-group.** *Int J Biometeorol* 2006, **50:**342–348. doi:10.1007/s00484-006-0033-z.
50. Kovats RS, Hajat S, Wilkinson P: **Contrasting patterns of mortality and hospital admissions during hot weather and heat waves in Greater London, UK.** *Occup Environ Med* 2004, **61:**898–898.
51. Morgan G, Sheppeard V, Khalaj B, Ayyar A, Lincoln D, Jalaludin B, *et al*: **The effects of bushfire smoke on daily mortality and hospital admissions in Sydney, Australia, 1994 to 2002.** *Epidemiology* 2010, **21:**47–55.
52. Wang XY, Barnett AG, Yu W, FitzGerald G, Tippett V, Aitken P, Neville G, McRae D, Verrall K, Tong S: *The impact of heatwaves on mortality and emergency hospital admissions from non-external causes in Brisbane*. Occ Env Med Online: Australia; 2010. doi:10.1136/oem.2010.062141.
53. Simpson R, Williams G, Petroeschevsky A, *et al*: **The short-term effects of air pollution on hospital admissions in four Australian cities.** *Aust NZ J Public Health* 2005, **29:**213–221.
54. Simpson R, Williams G, Petroeschevsky A, *et al*: **The short-term effects of air pollution on mortality in four Australian cities.** *Aust N Z J Public Health* 2005, **29:**205–212.
55. Knowlton K, Rosenthal JE, Hogrefe C, Lynn B, Gaffin S, Goldberg R, *et al*: **Assessing ozone-related health impacts under a changing climate.** *Environ Health Perspect* 2004, **112:**1557–1563.
56. Derwent RG: **Atmospheric chemistry**. In *Air Pollution and Health*. Edited by Holgate S, Samet J, Koren H. New York: Academic Press; 1999.
57. Ren C, Williams GM, Morawska L, Mengersen K, Tong S: **Ozone modifies associations between temperature and cardiovascular mortality: analysis of the NMMAPS data.** *Occup Environ Med* 2008, **65:**255–260.
58. Schwartz J, Spix C, Touloumi G, *et al*: **Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions.** *J Epidemiol Community Health* 1996, **50:**S3–S11.
59. Filleul L, Cassadou S, Médina S, Fabres P, Lefranc A, Eilstein D, Le Tertre A, Pascal L, Chardon B, Blanchard M, *et al*: **The relation between temperature, ozone, and mortality in nine French cities during the heat wave of 2003.** *Environ Health Perspect* 2006, **114:**1344–1347.
60. Storr J, Lenney W: **School holidays and admissions with asthma.** *Arch Dis Child* 1989, **64:**103–107.
61. Lincoln D, Morgan G, Sheppeard V, Jalaludin B, Corbett S, Beard J: **Return to school after term holidays is associated with increased hospital admissions for childhood asthma in Sydney, Australia.** *Public Health* 2006, **120:**854–862.
62. Bolger C, Tufvesson E, Anderson SD, *et al*: **The effect of inspired air conditions on exercise-induced bronchoconstriction and urinary CC16 levels in athletes.** *J Appl Physiol* 2011, **111:**1059–1065.
63. Guo Y, Barnett AG, Zhang Y, Tong S, Yu W, Pan X: **The short-term effect of air pollution on cardiovascular mortality in Tianjin, China: comparison of time series and case–crossover analyses.** *Sci Total Environ* 2010, **2010**(409)**:**300–306.

64. Rothman KJ: **No adjustments are needed for multiple comparisons.** *Epidemiology* 1990, **1**:43–46.
65. Australian Institute of Health and Welfare: **Ageing and aged care in Australia**. In *July 2008*. Canberra, Australia: Commonwealth of Australia; 2008:1–45.
66. Wilson L, Black DA, Veitch C: **Heatwaves and the elderly: the role of the GP in reducing morbidity.** *Aust Fam Physician* 2011, **40**:637–40.
67. Astrom DO, Forsberg B, Rocklov J: **Heat wave impact on morbidity and mortality in the elderly population: a review of recent studies.** *Maturitas* 2011, **69**:99–105.
68. NSW Ministry of Health (2012): **Chronic care management program.** 2012. http://www0.health.nsw.gov.au/cdm/severe_chronic_disease_management_program.asp.

doi:10.1186/1476-069X-12-98
Cite this article as: Wilson *et al.*: The impact of heat on mortality and morbidity in the Greater Metropolitan Sydney Region: a case crossover analysis. *Environmental Health* 2013 **12**:98.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

4.7 Paper 8: The effect on mortality of an intervention to reduce air pollution from woodheater smoke

Title: Johnston, F.H., Hanigan, I.C., Henderson, S.B., and Morgan, G.G. (2013). Evaluation of interventions to reduce air pollution from biomass smoke on mortality in Launceston, Australia: retrospective analysis of daily mortality, 1994-2007. BMJ, 346, e8446. doi:10.1136/bmj.e8446

Stage of publication: Published peer-reviewed paper

Journal Impact Factor: 17.4

My contribution: In this paper I worked with my co-authors to design the analysis based on the hypothesis proposed by Dr Johnston and Prof. Morgan. I took responsibility for the data processing stage to create the exposure data and cleaned the health outcome data. I was responsible for the decision to use age-adjusted Poisson regression of death counts rather than the alternative to use Standardised Mortality Ratios on the basis of my extensive review of the literature about these approaches when comparing multiple study populations. I conducted all data analyses. I produced figures and summarised the results. I helped interpret and write about the results and discussion. I also assisted prepare the text for the methods section of the paper. I prepared the study proposal for ethics committee review for our access to the restricted mortality data.

Approvals required for inclusion of Paper 8 in this thesis:

Name	Notes	Signature/date
Dr. Fay H. Johnston		2015-08-21
Mr. Ivan Hanigan		2015-09-28
Dr. Sarah B. Henderson		2015-10-06
Prof. Geoffrey G. Morgan		2015-11-15
Copyright	This is an open-access article distributed under the terms of the CC BY NC licence.	2015-10-07
Ethics Approval	ANU Human Research Ethics Protocol: 2008/199. Date entered 26/04/2008, variation = 03/06/2011 (extension of study period, addition of more data).	2008-04-26

BMJ

BMJ 2013;345:e8446 doi: 10.1136/bmj.e8446 (Published 8 January 2013) Page 1 of 11

RESEARCH

Evaluation of interventions to reduce air pollution from biomass smoke on mortality in Launceston, Australia: retrospective analysis of daily mortality, 1994-2007

OPEN ACCESS

Fay H Johnston *research fellow*[1], Ivan C Hanigan *research associate*[2 3], Sarah B Henderson *epidemiologist*[4], Geoffrey G Morgan *associate professor*[5 6]

[1]Menzies Research Institute Tasmania and Rural Clinical School, University of Tasmania, Tasmania, Australia; [2]National Centre for Epidemiology and Population Health, Australian National University ; [3]School of Plant Science, University of Tasmania, Tasmania, Australia; [4]British Colombia Centre for Disease Control, Vancouver, Canada; [5]University Centre for Rural Health-North Coast, University of Sydney, Sydney, Australia; [6]Northern New South Wales Local Health District, New South Wales, Australia

Abstract

Objective To assess the effect of reductions in air pollution from biomass smoke on daily mortality.

Design Age stratified time series analysis of daily mortality with Poisson regression models adjusted for the effects of temperature, humidity, day of week, respiratory epidemics, and secular mortality trends, applied to an intervention and control community.

Setting Central Launceston, Australia, a town in which coordinated strategies were implemented to reduce pollution from wood smoke and central Hobart, a comparable city in which there were no specific air quality interventions.

Participants 67 000 residents of central Launceston and 148 000 residents of central Hobart (at 2001 census).

Interventions Community education campaigns, enforcement of environmental regulations, and a wood heater replacement programme to reduce ambient pollution from residential wood stoves started in the winter of 2001.

Main outcome measures Changes in daily all cause, cardiovascular, and respiratory mortality during the 6.5 year periods before and after June 2001 in Launceston and Hobart.

Results Mean daily wintertime concentration of PM_{10} (particulate matter with particle size <10 µm diameter) fell from 44 µg/m^3 during 1994-2000 to 27 µg/m^3 during 2001-07 in Launceston. The period of improved air quality was associated with small non-significant reductions in annual mortality. In males the observed reductions in annual mortality were larger and significant for all cause (−11.4%, 95% confidence interval −19.2% to −2.9%; P=0.01), cardiovascular (−17.9%, −30.6% to −2.8%; P=0.02), and respiratory (−22.8%, −40.6% to 0.3%; P=0.05) mortality. In wintertime reductions in cardiovascular (−19.6%, −36.3% to 1.5%; P=0.06) and respiratory (−27.9%, −49.5% to 3.1%; P=0.07) mortality were of borderline significance (males and females combined). There were no significant changes in mortality in the control city of Hobart.

Conclusions Decreased air pollution from ambient biomass smoke was associated with reduced annual mortality in males and with reduced cardiovascular and respiratory mortality during winter months.

Introduction

Despite a vast amount of literature on the health effects of air pollution, few studies have investigated shifts in outcomes with public health interventions to improve ambient air quality. Previous examples of such studies include reduced smoke pollution in winter from biomass in a small community in Montana,[1] the temporary closure of a steel mill in Utah,[2-5] the reduction of the sulphur content of petrol in Hong Kong,[6] the cessation of coal sales in Dublin,[7] and the evaluation of more gradual improvements in urban air quality in Germany,[8] Switzerland,[9] and the United States.[10] These studies have reported decreases in a wide range of health outcomes associated with reductions in exposure to air pollution, including respiratory symptoms, mortality, and premature birth, and they have helped to quantify the health benefits of improved air quality. We assessed changes in mortality associated with an intervention to reduce ambient biomass smoke from domestic wood heaters.

Correspondence to: F H Johnston fay.johnston@utas.edu.au

Extra material supplied by the author (see http://www.bmj.com/content/345/bmj.e8446?tab=related#webextra)

Appendix 1: Sensitivity analysis showing influence of varying width of moving average used to smooth daily mortality data in reference population of Tasmania

Appendix 2: Analyses using single combined models with an interaction term for the period*city effect to statistically compare results for individual cities

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ 2013;345:e8446 doi: 10.1136/bmj.e8446 (Published 8 January 2013)

Biomass smoke is produced by the combustion of organic matter and includes emissions from domestic solid fuel use and landscape fires, with both indoor and outdoor sources contributing to the global burden of mortality.[11 12] While most research on the health effects of biomass smoke has concentrated on fine particulate matter (with particle size <2.5 µm diameter; $PM_{2.5}$) as the primary pollutant of concern, the complex smoke mixture contains numerous toxic co-pollutants, including volatile organic compounds and gases.[13] The toxicology and epidemiology of biomass smoke have been less well characterised than that of particulate emissions from industry and transportation.[14] Exposure to ambient biomass smoke, however, has been clearly associated with adverse respiratory outcomes and mortality,[14] and evidence is emerging for associations with adverse cardiovascular outcomes.[15 16] Intervention studies in Central America have replaced open cooking fires in dwellings with vented stoves to reduce indoor pollution from biomass smoke. Reported health improvements in the intervention groups include fewer respiratory symptoms,[17] lower blood pressure,[18] and babies with higher birth weights.[19] In Canada, an intervention using air filters to reduce indoor particulate matter from wood heaters found that biological markers of inflammation and endothelial dysfunction were reduced,[20] while in the US a 5 µg/m^3 reduction in outdoor fine particulate matter from biomass smoke was associated with reduced wheeze and respiratory infections in children.[1]

We assess the mortality changes associated with improvements to air quality after coordinated government interventions to reduce ambient air pollution from indoor wood stoves in the Tasmanian city of Launceston. We also use the same methods on a comparable population in the Tasmanian city of Hobart, which did not have any air quality interventions. We compared daily annual and wintertime all cause, cardiovascular, and respiratory mortality rates during a 6.5 year period of documented poor air quality in Launceston, with a 6.5 year period of improved air quality that followed the government funded interventions.

Methods

Setting

The Australian state of Tasmania (population 472 000 in 2001) is an island to the south of the continent, characterised by a colder and wetter climate than the rest of the Australia. Launceston (population 67 000 in 2001) is the second largest city in Tasmania and serves as the regional capital and tertiary medical referral centre for the northern part of the state. Hobart (population 148 000 in 2001) is the capital of Tasmania and the tertiary medical referral centre for the southern part of the state (fig 1⇓).

During the late 1980s and early 1990s, wood stoves became increasing popular for home heating throughout Tasmania. The impact on air quality was particularly severe in Launceston, which is in a river valley where both topographical and meteorological conditions limit atmospheric dispersion of air pollution (fig 2⇓).[21] During the 1990s, 66% of Launceston households used domestic wood stoves as the main source of heating,[22] and emissions from these accounted for an estimated 85% of particulate air pollution in winter.[23] The absolute number of wood heaters was estimated to be 17 500 in the year 2000.[24]

Pollution monitoring

In response to the perceived worsening of air quality in Launceston, a comprehensive monitoring campaign was started in 1991 to describe the spatial pattern of air pollution and to select the most appropriate site for long term monitoring of air quality. Simultaneous daily monitoring of concentrations of particulate matter less than 10 µm in diameter (PM_{10}) was conducted at five sites in 1991-93 with gravimetric sampling methods.[25] Measurements were found to be highly correlated with each other, except at one site to the east of the city where lower concentrations of particulate matter were generally recorded.[25] After these studies a representative site was selected for ongoing monitoring of the air quality of central Launceston and the 24 hour PM_{10} concentration was measured on a one in six day cycle from 1994 to 1997. Starting in May 1997 measurements were taken daily during winter, and starting in June 2001 measurements were taken daily all year. The study population was drawn from four statistical local areas (SLAs) within the city of Launceston for which the long term monitoring site was known to be representative.[25] We report PM_{10} because this was the only size fraction that was measured throughout the study period. Since 2001, however, concentrations of $PM_{2.5}$ have also been measured and these data show that the mean daily concentrations of $PM_{2.5}$ are highly correlated with the concentrations of PM_{10}, accounting for 50% of the annual average and 65% of the wintertime average.[26]

Interventions

The history of use of wood heaters, air pollution, and public health interventions in Launceston (fig 3⇓) has been fully documented elsewhere.[22 24] Local governments started responding to concerns about air quality in 1994 after a detailed air quality study conducted by an expert working group from 1991 to 1993.[25] These responses included the distribution of educational leaflets and the start of forecasts of wintertime air quality by the Bureau of Meteorology. They coincided with a state-wide marketing campaign by the local electricity supplier, promoting electricity as an affordable and non-polluting alternative (as most power in Tasmania is generated from hydroelectric power stations). During the 1990s there was a gradual decline in the proportion of homes heated primarily by wood throughout Tasmania and a gradual increase in the proportion heated by electricity.[27] While air pollution slowly began to improve during this time period,[22] substantive improvements in air quality did not occur until the start of coordinated government interventions in 2001. These included the Launceston Wood Heater Replacement Program, a $A2.05m (£1.34m, €1.66m) intervention funded through the Australian Natural Heritage Trust in July 2001 to June 2004.[22] This scheme greatly accelerated the general trend for home owners to replace wood with electricity as their primary source of domestic heating. By the end of the programme, the prevalence of wood stoves in Launceston had fallen from 66% to 30% of all households.[22] Follow-up surveys indicated that many wood heaters were decommissioned during this time without support from the government programme.[22] In addition to facilitating an overall reduction in the prevalence of wood heaters, the interventions also contributed to reducing pollution through improved operation of existing wood heaters by media advertising by the local council, school based education campaigns, and employment of environmental officers through the local council to monitor chimneys and provide targeted education, and, if necessary, infringement notices to home owners whose chimneys continued to emit excessive amounts of smoke after initial educational contact. A marked improvement in wintertime air quality coincided with the start of these interventions as illustrated by the photos in figure 2 and the PM_{10} data in figure 3.

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ 2013;345:e8446 doi: 10.1136/bmj.e8446 (Published 8 January 2013)

Meteorological data

We obtained data on daily weather from the Bureau of Meteorology for three weather stations in central Launceston and five weather stations in central Hobart.[28] For each city we calculated the average of the daily mean temperature and humidity inversely weighted by the distance of each weather station from the population weighted centre of the city.[29]

Population and outcome data

Population estimates were derived from the estimated resident population datasets available from the Australian Bureau of Statistics.[30] These were linearly interpolated by day between the estimated populations on 30 June of each census year (1991, 1996, 2001, and 2006) for the 14 year period from 1994 to 2007. Estimates were derived for Launceston, Hobart, and the rest of Tasmania.

Mortality data were obtained from Australian Bureau of Statistics. These data undergo considerable auditing for quality before being released for publication.[31 32] Causes of death were categorised according to the international classification of diseases (ICD) coding[33] into all cause (non-trauma), cardiovascular, and respiratory deaths (table 1⇓). Mortality data from January 1994 to November 2007 were included in the analysis. ICD coding changed from version 9 to version 10 during the study period, and we followed the mapping of the Australian National Casemix and Classification Centre to ensure that extracted codes included the same causes of death before and after the change in coding.[33] Hospital admissions data were provided by the Tasmania Department of Health and Human Services and were used to identify epidemics of respiratory infections by extracting the daily counts for admissions for pneumonia and flu (ICD9-9 480-487 or 507; ICD-10 J10-18 and J69-70).

Statistical analyses

We calculated annual concentrations of PM_{10} with one in six day sampling for the 1994-2000 period and daily monitoring data for the 2001-07 period. Winter (June, July, and August) PM_{10} concentrations were calculated from daily monitoring data from 1997. We used *t* tests to compare mean annual and wintertime concentrations from before and after the intervention periods.

Daily direct age standardised mortality rates for all of Tasmania, Launceston, and Hobart, were calculated and aggregated to describe the annual and wintertime mortality rates in the periods before and after the intervention (table 2⇓).

The effect of the intervention on all cause, cardiovascular, and respiratory mortality was assessed with age stratified time series Poisson regression models. The periods before and after the intervention were included as an indicator variable, and the interpolated age specific populations were used as an offset.[34] All models controlled for the effects of meteorology, epidemics of respiratory infections, and secular trends in daily mortality in Tasmania with the following equation:

$$\log(O_{ij}) = \text{period} + \text{age}_j + T_i + T_{i-2} + H_i + H_{i-2} + \text{DOW}_i + \text{flu}_i + \text{SecularTrend} + \text{offset}(\log(\text{Pop}_{ij}))$$

where: O_{ij} = observed number of cases on day_i in age_j; period = indicator variable for periods before and after the intervention; age = age groups by 15 year intervals to age 74, then 5 year intervals to 84, and ages ≥85; T_i = daily mean temperature (°C); T_{i-2} = average of three day lagged temperature day_i (°C); H_i = daily mean dew point (°C); H_{i-2} = average of three day lagged dew point (°C); flu = indicator variable for days when the 14 day moving average of hospital admissions for flu and pneumonia in Tasmania exceeded the 95th centile; SecularTrend = the 150 day moving average of the daily directly standardised cause specific mortality rates for Tasmania (excluding the intervention population of Launceston); and Pop_{ij} = interpolated population on day_i in age_j.

We conducted an identical statistical analysis for the control city of Hobart, after excluding its population from the data used to calculate secular mortality trends in Tasmania. We also conducted a subgroup analysis by sex.

We tested the sensitivity of our results to the use of different smoothing windows to calculate the state-wide secular trends in mortality; the definition of epidemics of respiratory infections (based on deaths from respiratory infections rather than hospital admissions for pneumonia or influenza); the use of indirect and direct age standardised mortality rates in the Poisson regression models (instead of age stratified regression); the inclusion of the winter of 2001 in the period before the intervention rather than the period after the intervention; and the use of the population of Northern Tasmania (where Launceston is situated) rather than all of Tasmania as the reference population for secular mortality trends.

Results

Air quality changed significantly over the study period (fig 3⇓). The annual mean concentration of PM_{10} before the intervention was 23.7 μg/m^3, based on the one in six day measurement cycle in 1997-2000 (n=237). This was higher (P<0.001) than the annual mean of 18.4 μg/m^3 after the intervention, based on the average of daily measurements in 2001-07 (n=2039). There was also a significant (P<0.001) decrease in the wintertime mean concentration of PM_{10} from 43.6 μg/m^3 before the intervention period to 27.0 μg/m^3 after the intervention (fig 3).⇓

The mean annual number of all cause deaths in Launceston was 577, 42% (243) of which were cardiovascular and 9% (54) of which were respiratory (table 1).⇓

There was a general trend towards reduced all cause, cardiovascular, and respiratory mortality throughout Tasmania during the study period, with the greatest magnitude observed for cardiovascular mortality (table 2). The reduction in respiratory mortality rates during winter months was larger in Launceston than in Hobart and the rest of Tasmania (table 2).⇓

We assessed the changes in mortality associated with the period of improved air quality (1 July 2001 to 30 December 2007) for the entire year and for winter months only after adjusting for the influence of daily meteorology, respiratory epidemics, population age structure, and secular mortality trends (table 3⇓). In the intervention community of Launceston, the period of improved air quality was associated with non-significant reductions in cardiovascular and respiratory mortality, which were not generally observed in the non-intervention community city of Hobart. In males, however the reductions in mortality in Launceston were large and significant for all cause (−11.4%, 95% confidence interval −19.2% to −2.9%; P=0.01), cardiovascular (−17.9%, −30.6% to −2.8%; P=0.02), and respiratory (−22.8%, −40.6% to 0.3%; P=0.05) mortality.

When we restricted analyses to the winter months, the reductions in cardiovascular (−19.6%, −36.3% to 1.5%; P=0.06) and respiratory (27.9%, −49.5% to 3.1%; P=0.07) mortality (males and females combined) were of borderline significance (table 3⇓). There were no significant mortality changes in the non-intervention community of Hobart (table 3).

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ 2013;345:e8446 doi: 10.1136/bmj.e8446 (Published 8 January 2013)

Sensitivity analyses

Our results were sensitive to the window of smoothing that we used to calculate the secular trends in mortality in the rest of Tasmania. We tested a range of smoothing intervals from a 30 day moving average increasing in units of 30 days to a 300 day moving average. When we applied moving average intervals of 120 days or fewer, the point estimates for the effect of the intervention corresponded to greater reductions in mortality (see appendix 1). At smoothing intervals of 120 days or longer, the point estimates were more stable but the width of the confidence intervals gradually increased. For this reason we have taken a conservative approach and report results from analyses using a moving average interval of 150 days.

These results were robust to different methods of demographic adjustment and the different definitions of respiratory epidemics. We found similar results when we repeated the analysis with the winter of 2001 included in the period before rather than in the period after the intervention. When we used data from Northern Tasmania instead of all of Tasmania to determine secular mortality trends, there was a slightly greater reduction in mortality associated with the period after the intervention in Launceston.

Discussion

Summary of main findings

Improved air quality after coordinated interventions in the study community of Launceston, Tasmania, was associated with reductions in all cause, cardiovascular, and respiratory mortality. This trend was greatest for wintertime cardiovascular and respiratory mortality. In sex specific analyses of annual mortality, associations were stronger and of greater magnitude in males.

Strengths and limitations of the study

Examples of population level environmental interventions are rare. The major strength of this study was our ability to investigate whether an intervention to reduce outdoor biomass smoke from domestic wood heaters was associated with a measurable reduction in mortality. Another strength was our ability to compare effects between the intervention population (Launceston) and a control population (Hobart). We report the results of independent statistical models for each city because this fulfilled our objective of assessing whether the intervention was associated with mortality changes in Launceston. Combined models using pooled data from both cities with an interaction term indicated that the differences between the two populations were significant for the reduction in annual cardiovascular mortality (males and females combined) and for annual all cause mortality in men (see appendix 2). An important limitation was the relatively small study population, which reduced the statistical power of our study. Although the magnitude of most associations was relatively large, confidence intervals were wide. This was especially the case for respiratory outcomes which comprised just 10% of all deaths.

Potential confounding factors

We were able to measure and adjust for known and measurable confounders such as age, temperature, humidity, and respiratory epidemics. It is challenging, however, to separate the influence of secular trends in mortality. We included smooth daily mortality data from all of Tasmania to adjust for secular trends because the entire state has similar distributions of health outcomes, socioeconomic status, and demographic structure.[35] The changing prevalence of population risk factors through time, such as smoking and diabetes, is likely to have been similar.[35]

Comparison with other studies

We found three similar intervention studies. In a comparable intervention in the community of Libby, Montana, older wood heaters were exchanged with new, less polluting models during in the winter of 2006-07. Concentrations of fine particulate matter ($PM_{2.5}$) fell from a mean of 27.2 $\mu g/m^3$ in the two winters before the intervention to 19.7 $\mu g/m^3$ for two winters after. The respiratory heath of children was assessed by repeated annual surveys of their parents. A reduction of 5 $\mu g/m^3$ in $PM_{2.5}$ was associated with a 27% (95% confidence interval 3% to 45%) reduction in wheeze and even larger reductions in respiratory infections, including flu (52%, 43% to 61%) and throat infections (45%, 29% to 58%).[1]

Other studies have evaluated changes in mortality after interventions to reduce the sulphur content of petrol in Hong Kong and after the ban of coal sales for domestic heating in Dublin. In Hong Kong there was a documented 45% decline in mean ambient concentration of sulphur dioxide over five years, with no significant change in the ambient concentration of particulate matter after the intervention.[6] The reported outcomes were a 2.1% (0.9% to 3.3%) decrease in all cause mortality, a 3.9% (6.2% to 1.6%) decrease in respiratory mortality, and a 2.0% (3.7% to 0.3%) decrease in cardiovascular mortality. These effects are small compared with our reported decrease in cardiovascular mortality, but they are more precise given the much larger population of Hong Kong compared with Launceston. As with our study, the observed decreases were higher during the colder months of the year. The Hong Kong study, however, did not include a non-intervention population to help disentangle any influence of long term trends in mortality. A second study evaluated the cessation of coal sales in Dublin, which resulted in 16% and 64% declines in wintertime concentrations of sulphur dioxide and black smoke, respectively. While associated declines in both respiratory and cardiovascular mortality were reported,[7] these data are currently being reanalysed.[36]

Further research

Our results contribute to the emerging evidence for an association between exposure to biomass smoke and cardiovascular mortality. One time series study reported association between daily cardiovascular mortality and ambient biomass smoke pollution from wood heaters[15] and another reported an association between biomass smoke from bushfires and out of hospital cardiac arrests.[37] Other studies, however, have had inconclusive results, and further research is needed.[38-40]

Conclusions and implications for policy

Observational studies are often the only practical option for evaluating the efficacy of population scale interventions. Adjustment for known and unknown confounding variables in observational studies is challenging. Given the clear and consistent difference in results between the intervention and the non-intervention populations, our findings suggest that the improved air quality in Launceston was associated with reductions in mortality. Furthermore, these results are consistent with the available literature concerning particulate air pollution.[41] Our findings highlight the potential for important public health gains from interventions to reduce ambient pollution from biomass smoke.

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ 2013;345:e8446 doi: 10.1136/bmj.e8446 (Published 8 January 2013)

What is already known on this topic

Though much research has shown the adverse effects of air pollution on health, few studies have investigated shifts in health outcomes associated with public health interventions to improve ambient air quality

A previous intervention study found that reduced outdoor wood smoke was associated with improvements in children's respiratory health

What this study adds

Coordinated interventions, including community education, law enforcement, and incentives to reduce pollution from domestic wood heaters, substantially improved winter air quality in Launceston, Tasmania

The period of improved air quality was associated with an overall trend towards reduced cardiovascular and respiratory mortality

We thank Douglas Dockery and Mark Clements for their support in development of the statistical methods. Grant Williamson assisted with preparation of the figures. James Markos provided the photographs used in figure 2.

Contributors: FHJ conceived the study, contributed to study design, and prepared the manuscript. ICH contributed to study design, conducted the analyses, and contributed to the manuscript. SBH and GGM contributed to study design and preparation of the manuscript. All authors read and approved the manuscript. All authors had full access to all of the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. FHJ and ICH are guarantors.

Funding: This study was supported by the Australian Research Council (LP0882048), the Tasmanian Government Department of Health and Human Services, and the Environment Protection Authority Tasmania. The funding agencies had no role in the study design, analysis, or preparation of the manuscript.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: This study was approved by the ANU human research ethics committee (2008/199) and the Tasmanian human research ethics committee (H0010047).

Data sharing: No further data are available.

1 Noonan CW, Ward TJ, Navidi W, Sheppard L. A rural community intervention targeting biomass combustion sources: effects on air quality and reporting of children's respiratory outcomes. *Occup Environ Med* 2012;69:354-60.
2 Parker JD, Mendola P, Woodruff TJ. Preterm birth after the Utah Valley steel mill closure: a natural experiment. *Epidemiology* 2008;19:820-23
3 Pope CA. Particulate pollution and health: a review of the Utah valley experience. *J Expo Anal Env Epid* 1996;6:23.
4 Pope CA. Respiratory hospital admissions associated with PM_{10} pollution in Utah, Salt Lake, and Cache Valleys. *Arch Environ Health* 1991;46:90-97.
5 Pope CA, Schwartz J, Ransom MR. Daily mortality and PM_{10} pollution in Utah Valley. *Arch Environ Health* 1992;47:211-17.
6 Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: an intervention study. *Lancet* 2002;360:1646-52.
7 Clancy L, Goodman P, Sinclair H, Dockery DW. Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study. *Lancet* 2002;360:1210-14.
8 Heinrich J, Hoelscher B, Frye C, Meyer I, Pitz M, Cyrys J, et al. Improved air quality in reunified Germany and decreases in respiratory symptoms. *Epidemiology* 2002;13:394-401.
9 Schindler C, Keidel D, Gerbase MW, Zemp E, Bettschart R, Brandli O, et al. Improvements in PM_{10}-exposure and reduced rates of respiratory symptoms in a cohort of Swiss adults (SAPALDIA-study). *Am J Respir Crit Care Med* 2009;179:579-87.
10 Pope CA, Ezzati M, Dockery DW. Fine-particulate air pollution and life expectancy in the United States. *N Engl J Med* 2009;360:376-86.
11 Ezzati M, Lopez A, Rodgers A, Vander Hoorn S, Murray C. Selected major risk factors and global and regional burden of disease. *Lancet* 2002;360:1347-60.
12 Johnston FH, Henderson SB, Chen Y, Randerson JT, Marlier M, DeFries RS, et al. Estimated global mortality attributable to smoke from landscape fires. *Environ Health Perspect* 2012;120:659-701.
13 Reid JS, Koppmann R, Eck TF, Eleuterio DP. A review of biomass burning emissions part II: intensive physical properties of biomass burning particles. *Atmos Chem Phys* 2005;5:799-825.
14 Naeher LP, Brauer M, Lipsett M, Zelikoff JT, Simpson CD, Koenig JQ, et al. Woodsmoke health effects: a review. *Inhal Toxicol* 2007;19:67-106.
15 Sanhueza PA, Torreblanca MA, Diaz-Robles LA, Schiappacasse LN, Silva MP, Astete TD. Particulate air pollution and health effects for cardiovascular and respiratory causes in Temuco, Chile: a wood-smoke-polluted urban area. *J Air Waste Manag Assoc* 2009;59:1481-8.
16 McGowan JA, Hider RN, Chacko E, Town GI. Particulate air pollution and hospital admissions in Christchurch, New Zealand. *Aust NZ J Publ Heal* 2002;26:23-9.
17 Romieu I, Riojas-Rodriguez H, Marron-Mares AT, Schilmann A, Perez-Padilla R, Masera O. Improved biomass stove intervention in rural Mexico: impact on the respiratory health of women. *Am J Respir Crit Care Med* 2009;180:649-56.
18 McCracken JP, Smith KR, Díaz A, Mittleman MA, Schwartz J. Chimney stove intervention to reduce long-term wood smoke exposure lowers blood pressure among Guatemalan women. *Environ Health Persp* 2007;115:996-1001.
19 Thompson LM, Bruce N, Eskenazi B, Diaz A, Pope D, Smith KR. Impact of reduced maternal exposures to woodsmoke from an introduced chimney stove on newborn birth weight in rural Guatemala. *Environ Health Persp* 2011;119:1489-94.
20 Allen RW, Carlsten C, Karlen B, Leckie S, van Eeden S, Vedal S, et al. An air filter intervention study of endothelial function among healthy adults in a woodsmoke-impacted community. *Am J Respir Crit Care Med* 2011;183:1222-30.
21 Keywood MD, Ayers GP, Gras JL, Gillett RW, Cohen DD. Size distribution and sources of aerosol in Launceston, Australia, during winter 1997. *J Air Waste Manag Assoc* 2000;50:418-27.
22 CSIRO Atmospheric Research. Woodheaters in Launceston—impacts on air quality. Department of the Environment and Heritage, 2005:52.
23 Department of Infrastructure Water and the Environment (DPIWE). Discussion paper on air quality management and policy development in Launceston. Tasmanian Department of Infrastructure Water and the Environment, 2000.
24 Power M. Air pollution dispersion within the Tamar Valley [thesis]. University of Tasmania, 2001.
25 Lyons L. Air pollution, environmental and respiratory diseases, Launceston and Upper Tamar Valley Tasmania. Launceston City Council, 1996;63.24.
26 Environment Protection Authority Tasmania. Tasmanian air quality strategy. Tasmanian Department of Tourism, Arts and the Environment, 2006.
27 Australian Bureau of Statistics. Energy use and conservation. 2011. www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4602.0.55.001Main+Features1Mar%202011?OpenDocument
28 National Climate Centre of the Bureau of Meteorology.Daily or three hourly weather data for bureau of meteorology stations. Bureau of Meteorology, 2010.
29 Hanigan IC, Hall G, Dear KB. A comparison of methods for calculating population exposure estimates of daily weather for health research. *Int J Health Geogr* 2006;5:38.
30 Australian Bureau of Statistics. Population estimates by age and sex, 2006. 2006. www.abs.gov.au/AUSSTATS/abs@.nsf/allprimarymainfeatures/2F33D3E91BA0F08FCA2574A900130EE7?opendocument
31 Australian Bureau of Statistics. Quality declaration deaths, Australia, 2010. Canberra, 2012. www.abs.gov.au/AUSSTATS/abs@.nsf/Latestproducts/3303.0Quality%20Declaration02010?opendocument&tabname=Notes&prodno=3303.0&issue=2010&num=&view
32 Australian Bureau of Statistics. Causes of death Australia 1964-2007 (3303.0). Health Statistics Unit, 2010. www.abs.gov.au/ausstats/abs@.nsf/mf/3303.0
33 National Casemix and Classification Centre. IDC-10 mapping tables. University of Wollongong, 2010.
34 Roalfe A, Holder R, Wilson S. Standardisation of rates using logistic regression: a comparison with the direct method. *BMC Health Serv Res* 2008;8:275.
35 Epidemiology Unit, Population Health. Health Indicators Tasmania. Department of Health and Human Services, 2008:124.
36 Dockery DW, Rich DQ, Goodman PG, Clancy L, Ohman-Strickland P, George P, et al. Effect of air pollution control on mortality and hospital admissions in Ireland. Health Effects Institute Annual Conference, Chicago, Illinois, April 2012. www.healtheffects.org/Pubs/AnnualConference2012Program&Abstracts.pdf
37 Dennekamp M, Erbas B, Sim M, Glass D, Keywood M, Abramson M, et al. Air pollution from bushfires and out of hospital cardiac arrests in Melbourne Australia. *Epidemiology* 2011;22:S53.
38 Johnston FH, Hanigan IC, Henderson SB, Morgan GG, Bowman DMJS. Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994-2007. *Environ Res* 2011;111:811-6.
39 Hales S, Salmond C, Town GI, Kjellstrom T, Woodward A. Daily mortality in relation to weather and air pollution in Christchurch, New Zealand. *Aust NZ J Publ Heal* 2000;24:89-91.
40 Morgan G, Sheppeard V, Khalaj B, Ayyar A, Lincoln D, Jalaludin B, et al. The effects of bushfire smoke on daily mortality and hospital admissions in Sydney, Australia, 1994 to 2002. *Epidemiology* 2010;21:47-55.
41 Zanobetti A, Schwartz J. The effect of fine and coarse particulate air pollution on mortality: a national analysis. *Environ Health Persp* 2009;117:898.

Accepted: 03 December 2012

Cite this as: *BMJ* 2012;345:e8446

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the license. See: http://creativecommons.org/licenses/by-nc/2.0/ and http://creativecommons.org/licenses/by-nc/2.0/legalcode.

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ 2013;345:e8446 doi: 10.1136/bmj.e8446 (Published 8 January 2013)

Tables

Table 1| Mean number of deaths for all cause, cardiovascular, and respiratory mortality in Launceston, Tasmania, 1994-2007

Mortality	ICD codes*	Annual	Winter
All cause	ICD-9 <800; ICD-10 A00-R99	577	158
Cardiovascular	ICD-9 390-459; ICD-10 I00-I99 (excl I67.3, I68, I88, I97.8, I97.9, I98), G45 (excl G45.3), G46, M30, M31, R58	243	68
Respiratory	ICD-9 codes 460-519; ICD-10 J00-J99 (excl J95.4-J95.9), R09.1, R09.8	54	17

*ICD (international classification of diseases) coding changed from version 9 to version 10 during study period. We followed mapping of National Casemix and Classification Centre to ensure that extracted codes included same causes of death before and after change in coding.[33]

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ 2013;345:e8446 doi: 10.1136/bmj.e8446 (Published 8 January 2013)

Table 2| **Annual and wintertime age standardised mortality in Tasmania, Launceston, and Hobart during calendar years 1994-2000 and 2001-07, before and after intervention to improve air quality in Launceston**

	Deaths per 1000 person years		
	1994-2000	2001-07	Percent decrease (95% CI)
All Tasmania			
All cause mortality:			
All year	8.33	7.12	14.6 (5.2 to 24.0)
Winter	9.32	7.88	15.5 (6.4 to 24.6)
Cardiovascular mortality:			
All year	3.73	2.65	28.9 (5.6 to 42.2)
Winter	4.28	2.96	31.1 (18.6 to 43.7)
Respiratory mortality:			
All year	0.78	0.62	20.5 (8.3 to 32.7)
Winter	1.00	0.76	22.9 (7.6 to 38.3)
Launceston			
All cause mortality:			
All year	8.57	7.42	13.4 (2.9 to 24.0)
Winter	9.20	8.08	16.0 (3.9 to 28.2)
Cardiovascular mortality:			
All year	3.88	2.74	29.5 (14.0 to 45.1)
Winter	4.52	2.96	34.4 (22.5 to 6.4)
Respiratory mortality:			
All year	0.86	0.64	24.6 (4.7 to 44.6)
Winter	1.16	0.76	33.0 (4.6 to 61.4)
Hobart			
All cause mortality:			
All year	8.25	7.22	12.5 (3.4 to 21.5)
Winter	9.52	8.12	14.7 (5.1 to 24.2)
Cardiovascular mortality:			
All year	3.58	2.68	25.2 (12.6 to 37.7)
Winter	4.16	2.96	28.6 (14.7 to 42.5)
Respiratory mortality:			
All year	0.76	0.64	15.7 (0.5 to 30.9)
Winter	1.00	0.88	12.3 (−13.4 to 7.9)

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ 2013;345:e8446 doi: 10.1136/bmj.e8446 (Published 8 January 2013)

Table 3| **Percentage change* in all cause, cardiovascular, and respiratory mortality in Launceston and Hobart, Tasmania, from January 1994-May 2001 to June 2001-November 2007. Years 2001-07 correspond with period of improved air quality after series of coordinated interventions in Launceston**

	Launceston (intervention)		Hobart (control)	
	Percent change (95% CI)	P value	Percent change (95% CI)	P value
All year—males and females combined				
All cause mortality	−2.7 (−8.7 to 3.7)	0.40	1.4 (−3.0 to 6.0)	0.54
Cardiovascular mortality	−4.9 (−15.5 to 7.0)	0.40	0.9 (−7.1 to 9.6)	0.83
Respiratory mortality	−8.5 (−23.2 to 9.0)	0.32	4.8 (−7.4 to 18.6)	0.50
All year—males				
All cause mortality	−11.4 (−19.2 to −2.9)	0.01	0.7 (−5.4 to 7.2)	0.82
Cardiovascular mortality	−17.9 (−30.6 to −2.8)	0.02	−7.1 (−16.8 to 3.8)	0.19
Respiratory mortality	−22.8 (−40.6 to 0.3)	0.05	3.4 (−13.1 to 24.4)	0.67
All year—females				
All cause mortality	2.7 (−5.3 to 11.4)	0.52	−0.7 (−6.3 to 5.2)	0.80
Cardiovascular mortality	2.3 (−12.2 to 19.3)	0.77	3.6 (−7.6 to 16.2)	0.54
Respiratory mortality	1.0 (−18.9 to 24.4)	0.96	−1.4 (−15.5 to 15.1)	0.86
Wintertime—males and females combined				
All cause mortality	2.2 (−14.1 to 11.3)	0.73	−2.0 (−10.2 to 6.9)	0.64
Cardiovascular mortality	−19.6 (−36.3 to 1.5)	0.06	−7.0 (−20.8 to 9.2)	0.38
Respiratory mortality	−27.9 (−49.5 to 3.1)	0.07	8.0 (−16.9 to 40.4)	0.60

*Adjusted for age structure, meteorological conditions, and secular mortality trends in Tasmania.

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ 2013;345:e8446 doi: 10.1136/bmj.e8446 (Published 8 January 2013)

Figures

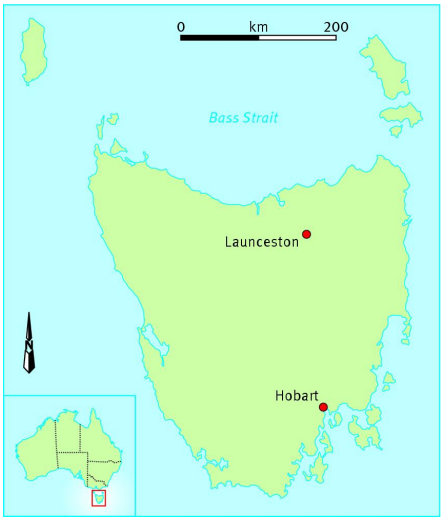


Fig 1 Map of Tasmania showing location of Launceston (intervention city) and Hobart (control)

Fig 2 Launceston, Tasmania, showing reduced visibility associated with smoke from domestic wood heaters (left) and same view on clear day (right)

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

BMJ 2013;345:e8446 doi: 10.1136/bmj.e8446 (Published 8 January 2013)

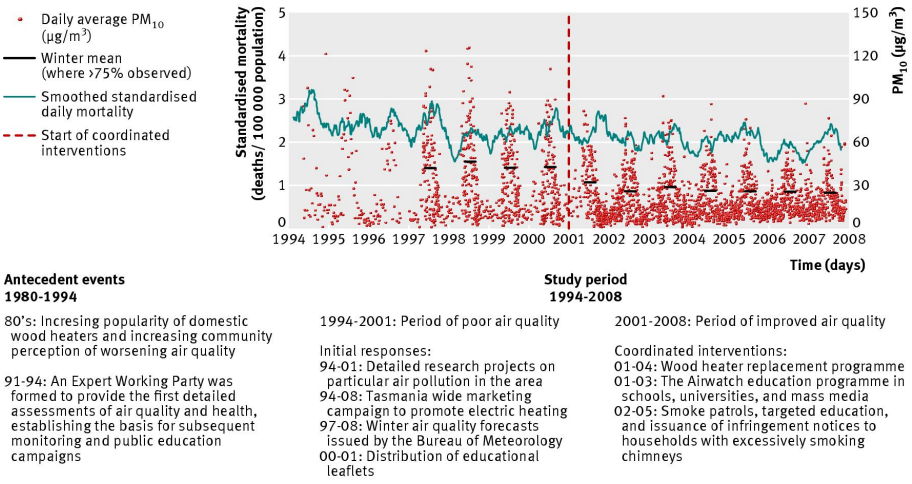


Fig 3 Air quality interventions, air quality data, and directly aged standardised mortality in Launceston, Tasmania 1991-2007

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe

Chapter 5

The effect on infectious diseases of local environmental factors

5.1 Paper 9: The effect on community gastroenteritis of weather

Title: Hall, G. V, Hanigan, I. C., Dear, K. B. G., and Vally, H. (2011). The influence of weather on community gastroenteritis in Australia. Epidemiology and Infection, 139(6), 927936. doi:10.1017/S0950268810001901

Stage of publication: Published peer-reviewed paper

Journal Impact Factor: 2.5

My contribution: I worked with my co-authors to design the analysis based on the hypothesis proposed by Prof. Hall and Dr. Vally. I took responsibility for the data processing stage to create the spatiotemporal exposure data, and I conducted all data analyses. In the first stage of preparing the exposure data I contributed the method of population weighted averages of weather for small areas based on the high resolution data from the population census collector districts, which I had previously developed and published (Hanigan et al 2006). For this paper I created the categorisation of postcodes into the different climatic zones and re-classified the climate zones based on the theories we proposed regarding our health outcome data (data available on request). During the analysis phase I worked with Prof. Dear to implement the statistical method for constrained polynomial distributed lag modelling. I was responsible for running the many data analysis experiments to systematically explore all potential combinations of variables, interactions and lag lengths. I contributed the idea for using the Akaike Information Criterion (AIC) to determine the optimal model structure. I produced all the figures and summarised the results. I helped to interpret and write about the results. I also prepared the text for the methods section of the paper. I was corresponding author and produced additional data analysis in response to requests by a reviewer. I was responsible for archiving the data and code, and subsequently published the exposure data to accompany this paper (dataset 4 of this thesis).

Approvals required for inclusion of Paper 9 in this thesis:

Name	Notes	Signature/date
Dr. Gillian V. Hall		2015-08-15
Mr. Ivan Hanigan		2015-09-28
Prof. Keith B. G. Dear		2015-10-29
Dr. Hassan Vally		2015-08-19
Copyright	License Number: 3723610175764 License date: Oct 07, 2015 Licensed content publisher: Cambridge University Press	2015-10-07
Ethics Approval	Our access to the gastroenteritis survey was approved by the Australian Department of Health and Ageing.	

Epidemiol. Infect., Page 1 of 10. © Cambridge University Press 2010
doi:10.1017/S0950268810001901

The influence of weather on community gastroenteritis in Australia

G. V. HALL[1,2], I. C. HANIGAN[1]*, K. B. G. DEAR[1] AND H. VALLY[1]

[1] *National Centre for Epidemiology and Population Health, The Australian National University, ACT, Australia*
[2] *Medical School, The Australian National University, ACT, Australia*

(*Accepted 15 July 2010*)

SUMMARY

Infectious gastroenteritis is a common illness in Australia as elsewhere. Data from a year-long national gastroenteritis survey in 2001–2002 showed that gastroenteritis was more common in the northern and hotter part of Australia. These data were used to quantify associations between local weather variables and gastroenteritis in people aged >5 years while controlling for socioeconomic status. A distributed lag model was used to examine the influence of weather over a period of days prior to an event and the maximal effect was found at a lag of 2–5 days. The total effect over the preceding week indicated a relative increase from baseline in the probability of gastroenteritis of 2·48% (95% CI 1·01–3·97) for each degree rise (°C) over that period. Given the very high burden of gastroenteritis, this represents a substantial effect at the population level and has relevance for health predictions due to climate change.

Key words: Climate (impact of), community epidemics, gastroenteritis.

INTRODUCTION

Gastroenteritis due to infection remains a common illness in both underdeveloped and developed countries. Children are especially vulnerable and the rate of gastroenteritis in those aged <5 years in underdeveloped countries is estimated at 3·2 episodes per child-year [1]. Children aged 6–12 months have the highest rate of disease with a median of 4·8 episodes per child per year. The consequences of gastroenteritis can be serious, especially in young children and the elderly, with infectious diarrhoea accounting for over two million deaths each year [2]. The illness is usually well managed in developed countries and deaths are uncommon but gastroenteritis still causes a considerable burden in terms of lost productivity and cost [3, 4]. The estimated incidence of all infectious gastroenteritis in USA, Canada and Australia is around one case per person per year [5].

The cause of gastroenteritis is unknown in most people even in developed countries. In cases where a pathogen is found, the most commonly identified pathogens include bacteria (e.g. *Salmonella* and *Campylobacter*), viruses (e.g. rotavirus and norovirus) and parasites (e.g. *Giardia*, *Cryptosporidium*). Many pathogens show seasonality of infection but the pattern varies. Peak rates have been noted in winter (rotavirus), summer (*Salmonella*) or spring (*Campylobacter*) [6–8]. *Salmonella* and *Campylobacter* rates are higher in younger children although they occur fairly frequently at all ages [9]. The pathogen mix varies by age group with some pathogens like rotavirus making up most of known pathogens in young children, while they form only a small proportion of

* Author for correspondence: Mr I. C. Hanigan, National Centre for Epidemiology and Population Health, Building 62, The Australian National University, Acton, ACT 0200, Australia. (Email: Ivan.hanigan@anu.edu.au)

known cases in adults [9]. Studies in different parts of the world have found a relationship between various types of gastroenteritis and underlying seasonal temperature and rainfall changes [6, 7, 10–14]. However, the pathways underlying variation in seasonal rates are not well understood.

Gastroenteritis is caused by transmission of pathogens from other people, animals, water or food, with the proportion of illness occurring via each route varying across pathogens. Local weather conditions can potentially influence the risk of gastroenteritis in humans by affecting proliferation and carriage of pathogens in the local environment – in locally produced or contaminated food or water, or directly from animals or the general environs [15]. Weather can also affect human behaviour, which in turn can influence the risk of acquiring gastroenteritis, such as through increased exposure to animal droppings by camping or eating outside [16] which is more likely in warmer weather. Other seasonal factors apart from weather can also impact on risk of gastroenteritis, such as eating particular foods or engaging in certain activities at holiday times that lead to increase risk through increased exposure to pathogens.

In Australia as elsewhere, gastroenteritis is a common problem and there are an estimated 17 million cases of gastroenteritis annually (defined as three loose stools or two vomits in 24 h) [4]. Australia spans a wide range of latitudes and there is variation in weather patterns from hot and wet in the north, to dry inland, and to cooler temperatures in southern regions. Both seasonal patterns and a latitudinal gradient of rates of illness have been noted for certain pathogens. For example, mandatory notifications of *Salmonella* show that it is more common in late summer and in the hotter Northern Territory [17]. However, factors other than weather can also vary across the country. There is a different demographic pattern in the most northern areas where a greater proportion of the population are Indigenous (over 25% in Northern Territory) compared to about 2% in the country as a whole [18]. In addition, socioeconomic status and community facilities for sewerage and water supply may be poor in some northern areas, which may confound any apparent correlation between weather and infectious diarrhoeal disease.

The national survey of gastroenteritis in Australia in 2001–2002 found that overall there was a higher incidence of gastroenteritis in summer across the whole country, and in the Northern Territory across the whole year, suggesting that spatial and temporal weather patterns may be a causal factor [19]. The survey also collected information on socioeconomic status. In this study we use data collected from this survey to explore the relationship between gastroenteritis and localized spatio-temporal variations in temperature and rainfall in Australia, while controlling for socioeconomic status. The objective was to quantify any relationship between localized climatic factors and diarrhoeal illness.

METHODS

Survey design

Data on gastroenteritis were acquired through a national computer-assisted telephone survey conducted from September 2001 to August 2002. A stratified random sample was drawn of over 800 households in each state. The spatial distribution of the households is shown by postal area in Figure 1. There were more households sampled in densely populated capital cities in each state with lower numbers in rural areas since the sample was a random selection in each state. The size of the postal areas varies considerably with large areas in central regions and small areas in urban localities.

Households were selected by random digit dialling, and the respondent was selected to be the person in the household with the next birthday. The response rate for those households that were contacted and eligible was 68% and interviews were completed with 6087 respondents. Of these, 322 children aged <5 years were excluded for the analysis in this study as different pathogens and seasonality are to be expected in this group. Respondents were asked about episodes of being sick with vomiting or diarrhoea in the 4 weeks prior to interview. For those with chronic illness with diarrhoea, the episode had to be different from their usual pattern. The postcode of the household and date of onset of episodes were recorded. Information was also collected on age, sex, education level, household income, health insurance status (private or not), and Indigenous status. The case definition of gastroenteritis was three loose stools or two vomits in 24 h. There were 677 respondents who met this definition of gastroenteritis of whom 618 had complete data on postcode and date of onset. More details on the survey methodology and the other information that was collected are given elsewhere [20].

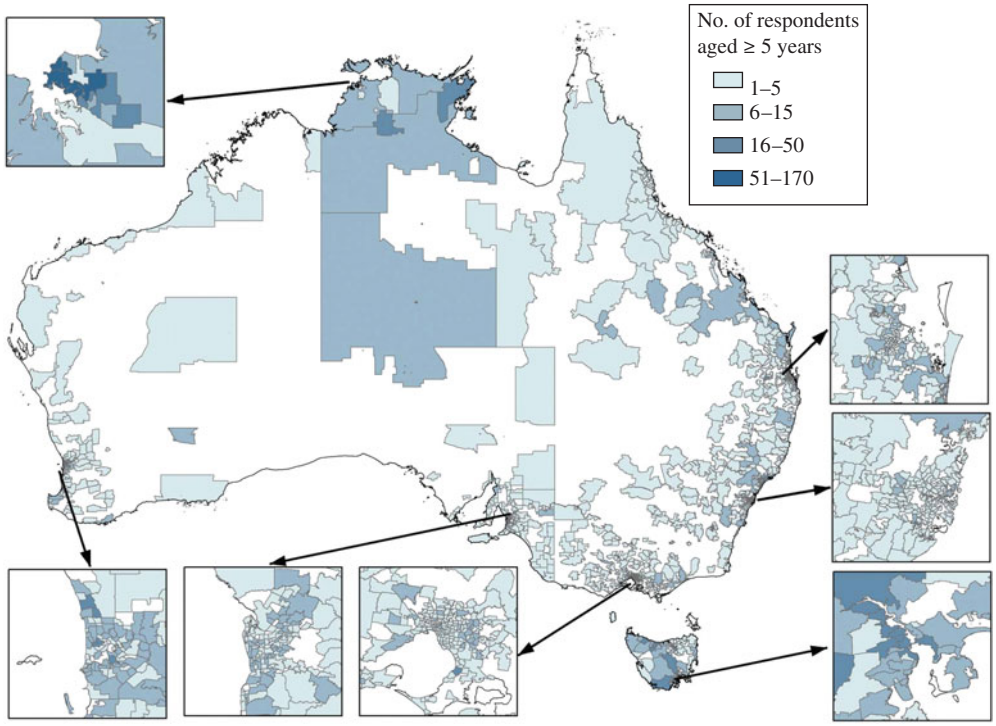


Fig. 1. Distribution of respondents by postal area.

Weather data

Weather data were obtained from the Bureau of Meteorology for average daily temperature, daily rainfall and relative humidity during the time of the gastroenteritis survey and for 90 days prior to commencement of the survey. Data were from all weather stations operating at the time that were within 50 km radius of the centroid of the Australian Bureau of Statistics 'census collectors' districts. The data were aggregated at the census collector's district level and subsequently at the postal area level (approximating postcodes), using a technique that weights data from different weather stations according to population density. Details of the methodology are given elsewhere [21].

The gastroenteritis survey data were arranged with each of the respondents contributing up to 28 days of data. For each day, the respondent either had no gastroenteritis, or had onset of gastroenteritis (first day of gastroenteritis). Where gastroenteritis occurred, the days after the onset of gastroenteritis were censored. Weather data were merged with the gastroenteritis survey data at the daily postal area level. Weather data were incomplete for a few respondents, and the final dataset contained 5670 respondents, 536 cases and about 158 000 person-days of data.

Descriptive seasonal and spatial analysis

To examine the spatio-temporal pattern of gastroenteritis over the observation period, data were examined by five temperature climatic zones. These were based on six zones described by the Bureau of Meteorology with two zones in Central Australia with hot dry summers coalesced into one [22]. The five zones are: hot humid summer; warm humid summer; hot dry summer; warm summer cold winter; mild warm summer cold winter. These zones are shown in Figure 2.

The proportion of respondents interviewed in that month who had gastroenteritis in the preceding 28 days was graphed in each of the five temperature climatic zones. The average monthly temperature, rainfall and relative humidity were then plotted on these graphs (Fig. 3).

Model for testing the association between local climatic factors and gastroenteritis

Associations between local ambient temperature, rainfall or relative humidity in the recent past and the development of gastroenteritis were examined using survival analysis (Cox proportional hazard model) in respondents aged >5 years. Each respondent

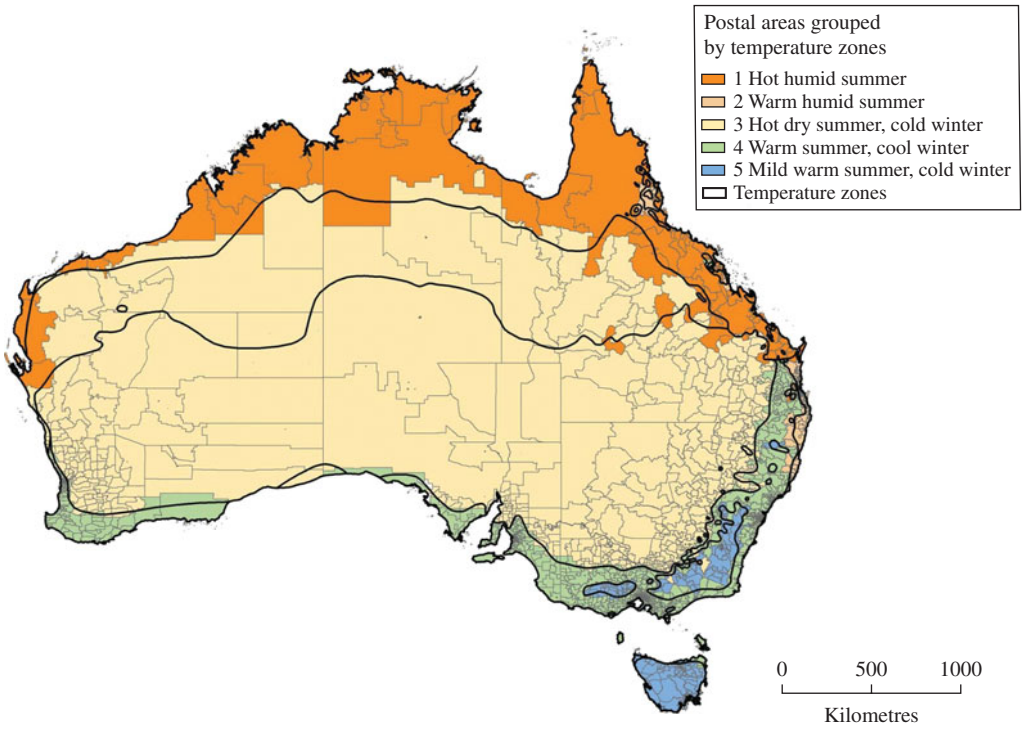


Fig. 2. Postal areas mapped to the six temperature zones in Australia (two zones in Central Australia were merged into one).

contributed 28 days of outcome data; subjects who had an episode were deemed to have 'failed' at the first day of onset, while subjects who had no episodes were censored at 28 days. Individual variables were tested for statistical significance in the models using Wald tests. When comparing models, the likelihood ratio test was used if the models were nested and Akaike's Information Criterion (AIC) was used to test comparative models containing different explanatory variables that were not nested. Statistical significance was set at $P<0{\cdot}05$ to assist with the model-building process.

To control for sociodemographic factors, an initial model was run containing the following variables: age, sex, education level (at least year-10 high school *vs.* less), household income (<$50 000/yr or more than), private health insurance (yes/no) and Indigenous status (yes/no). Statistically significant variables were then retained in all further models.

Categorical variables for climate zone and season were constructed in order to allow for effects due to season or zone other than temperature and other weather variables that might influence gastroenteritis.

Postal areas were mapped and aggregated to the six climatic zones based on their population weighted centre and the result is shown in Figure 2. The four seasons were based on the dates of the 4-week period of interest regarding onset of gastroenteritis (or not) for each respondent. The season for each day of analysis was categorized as spring (September–November), summer (December–February), autumn (March–May) or winter (June–August).

Distributed lag modelling

To assess the influence of local daily weather over time on the onset of gastroenteritis, time-varying covariates were constructed representing the weather in each postcode for up to 90 days preceding each of the 28 days of interest for each respondent. The effects of these lagged variables were assessed using polynomial distributed lag models, modified to impose continuity with the assumption of no effect beyond a chosen extent of lag effects [23]. These models allow a distinct effect at each lag, but constrain the effects to lie on a polynomial curve to avoid collinearity and overparameterizing the model [24]. Polynomials with 2, 3 and 4 degrees of freedom (D.F.) in each variable were explored as follows.

We allowed X_i to represent a variable X at lag i, for $i=0, 1, \ldots, 90$. Then instead of the model including 91 separate terms of the form $\beta_i X_i$, we constrained the

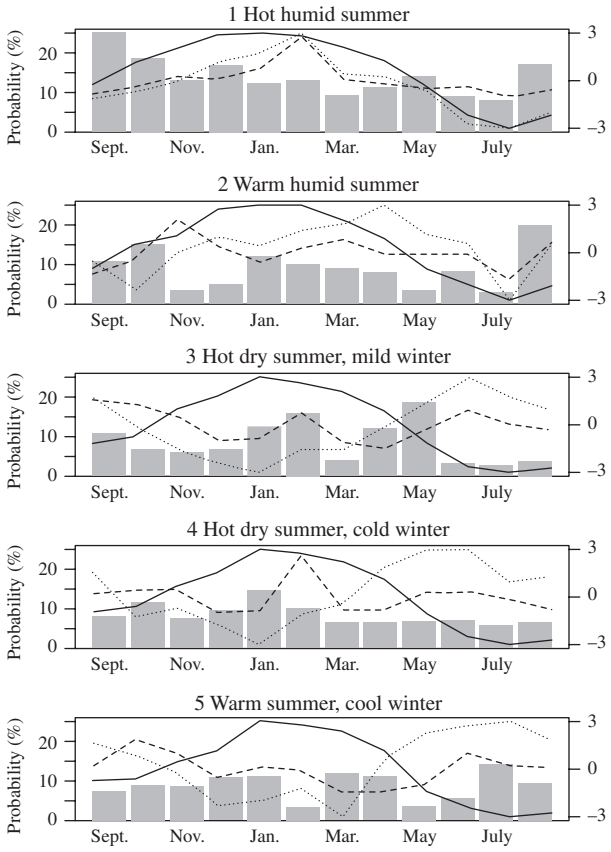


Fig. 3. Monthly percentage of respondents with gastroenteritis and average monthly temperature (—), rainfall (- - -) and relative humidity (·····) in the survey postal areas 2001–2002, by climatic zone (weather variables are plotted with normalized weather on the secondary y-axis as the magnitude changes between zones and parameter).

parameters β_i to lie on a polynomial of the form (for the 3 D.F. model):

$$\beta_i = \alpha_2(91-i)^2 + \alpha_3(91-i)^3 + \alpha_4(91-i)^4.$$

This polynomial was designed to have both value zero and slope zero at lag 91, 1 day beyond the maximum lag considered in the models. The model is fitted by including constructed covariates U_j to represent temperature:

$$U_j = \sum_{i=0}^{90} X_i\,(91-i)^j \quad (j=2,\ 3,\ 4).$$

We assessed the possibility that the temperature effect was curved at each lag within a constrained distributed lag model by modelling a 2–4 D.F. natural spline basis in place of linear terms in temperature X. We explored the constrained lag-modelling strategy by fitting a series of models, using a 2 D.F. spline of temperature in a 3 D.F. polynomial lag structure, constrained at weekly intervals up to 90 days, and then at each day near the optimum. We used AIC to select the optimal lag at which to constrain the effect estimate polynomial to zero. The choice of degrees of freedom of the spline of temperature was then checked in the constrained 3 D.F. polynomial lag structure by comparing models from spline of 4 D.F. to 1 D.F.

Selection of final model

While holding the statistically significant sociodemographic variables in all models, a series of models were explored to examine the relationship between temperature, zone and season with gastroenteritis. As well as accounting for possible influences relating to season or zone other than weather, the terms for climate zone and season also both contain information on temperature and other weather factors. A high degree of correlation is therefore to be expected between the local temperature variable and these categorical variables. Zone and season were first modelled separately from the temperature distributed lag polynomial and then together with the temperature polynomial, and the models were compared. The explanatory variables in each model are listed below:

Model 1: Season and zone,
Model 2: Temperature,
Model 3: Season, zone and temperature,
Model 4: Season and temperature,
Model 5: Zone and temperature.

Interactions were explored between the temperature polynomial and zone and season, and between temperature and sociodemographic variables. The variables for rain and relative humidity were explored in a similar way to the temperature variables.

Finally, variables that were statistically significant in any of the above five models were put into a combined model. Variables that remained statistically significant in the combined model were retained as the final model. The model fit of the largest possible model with all weather, zone and season variables was then compared with the final parsimonious model. Confidence intervals (CI) are at 95%. Analysis was done using Stata 9.2 [25].

RESULTS

Spatial and seasonal distribution of gastroenteritis

The probability of gastroenteritis in each temperature climatic zone is shown in Table 1. There is a higher

Table 1. *Probability (%) of gastroenteritis by temperature/humidity zone*

Climate zone	Cases (*n*)	Respondents (*n*)	Percentage gastroenteritis	Average temp. 1961–1990 (°C)	Average temp. in survey areas in 2001–2002 (°C)
1. Hot humid summer	116	853	13·6	26	25
2. Warm humid summer	46	515	8·9	20	21
3. Hot dry summer	38	442	8·6	21	18
4. Warm summer cold winter	259	3000	8·6	15	16
5. Mild warm summer cold winter	77	860	9·0	11	12
Total	536	5670	9·0	19	18

percentage of cases in the northern zone of Australia where the temperatures are higher and the summers are wet.

Figure 3 shows the monthly incidence of gastroenteritis for the period September 2001 to August 2002 in each of the five climatic zones along with monthly average temperature, relative humidity and rainfall.

Rainfall is higher in northern Australian (zone 1) and peaks in summer. By comparison, rainfall is more evenly distributed over the year in the other zones and is considerably less in central Australia (zone 3), eastern coastal areas (zone 4) and mountains in the south (zone 5). The northeastern coastal region (zone 2) has intermediate rainfall. Temperatures are higher in summer (December–February) in all zones. Overall temperatures are highest in the north (zone 1) at about 25 °C, decreasing as the zones become more southerly to about 12 °C (zone 5). Relative humidity shows most seasonality in the central region (zone 3) where it drops in summer. Other regions remain at around 70% for the whole year.

The probability of gastroenteritis in the previous 4 weeks was highest in northern Australia (zone 1) ranging from about 7% of respondents in winter to 25% in early spring (August–October) when temperatures start to rise. The probability of gastroenteritis did not continue to rise through the summer months. Zone 2 also had higher rates in spring while zones 3 and 4 had higher rates in summer (January). These descriptive statistics suggest that the potential relationship between weather and gastroenteritis may vary across climate zones.

Controlling for sociodemographic status

The dependence of risk on age in years was best modelled as a linear term plus a square-root-transformed term. Age, sex and health insurance were statistically significant, and were held in all further models used to assess the influence of weather on the outcome of gastroenteritis. Indigenous status, income and education level were no longer statistically significant when the other sociodemographic variables were in the model, and so were dropped.

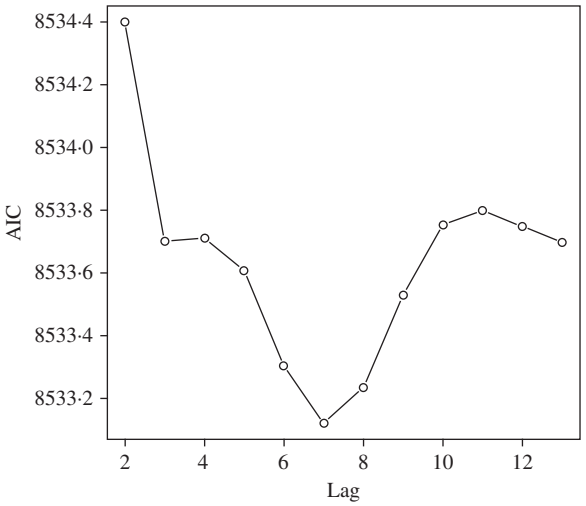


Fig. 4. Akaikes Information Criteria (AIC) for models constrained at 3–15 days. Optimal model found with maximum lag at day 7 (constrained to zero at day 8).

Distributed lag models

A statistically significant association was found between temperature as a distributed lag model with 3 D.F. and the outcome of gastroenteritis, while the variables for rain and relative humidity were not statistically significant in any form, or as any distributed lag polynomial. The model fit for temperature was best when the polynomial was 3 D.F. (AIC 8915·01 compared to 2 D.F., AIC 8918·37). The range of average daily temperatures included in the model was from −2 °C to 36 °C. In comparing the extent of the temperature-distributed lag function, constraining to zero at day 8 was optimal (Fig. 4).

In testing natural spline temperature response functions to assess if the temperature effect was curved at each lag, none fitted the data significantly better than a straight line model. The likelihood ratio test statistic comparing 2 D.F. with a straight line was 1·54 ($P=0·67$). The temperature effect was therefore modelled as a 3 D.F. polynomial distributed lag model constrained to zero at day 8, with a straight line temperature response function at each lag.

Models 1–5 were constructed while controlling for the sociodemographic variables for age, sex and health insurance. The influence of sociodemographic factors are explored more fully elsewhere [19]. When zone and season were in a model together (model 1) without temperature, both variables were statistically significant and AIC was 8918·78. When temperature was in a model without zone and season (model 2), this variable was highly statistically significant ($P<0·001$) and AIC was 8915·01. When all three variables were in a model together (model 3), none were significant suggesting a high degree of correlation between the variables, and AIC was 8921·30. When temperature was in a model with either season (model 4) or zone (model 5), temperature was significant and zone and season were not. Interaction terms were explored and no interaction terms involving season or zone gave a model with a better fit. The temperature model (model 2) had the lowest AIC indicating the best fit and was therefore adopted as the final parsimonious model with explanatory factors of age (linear and square root terms), sex, health insurance, and temperature as a 3 D.F. polynomial distributed lag model constrained to zero at day 8. The effect of temperature over the 8 days in this model is shown in Figure 5.

The effect at day zero (i.e. day of onset of gastroenteritis) is negative but the 95% CI there includes the value of no effect [hazard ratio (HR) of 1]. The temperature effect is highest on days 2–5 where the HR is about 1·01 (1% increase relative to baseline) per °C for any one day. The cumulative effect across the 8 days gives an HR of 1·0248 or an increase of 2·48% (95% CI 1·01–3·97) relative to the baseline probability of gastroenteritis. That is, for a 1 °C rise in temperature maintained for 8 days, there is a 2·48% increase relative to baseline in the probability of diarrhoea on the eighth day. For an increase in temperature of 5 °C, the relative risk (HR) is $(1·0248)^5$ or 1·13, an increase of 13% from the baseline probability of gastroenteritis. The same modelling process was repeated for rainfall and relative humidity. Neither of these variables were statistically significant in any model.

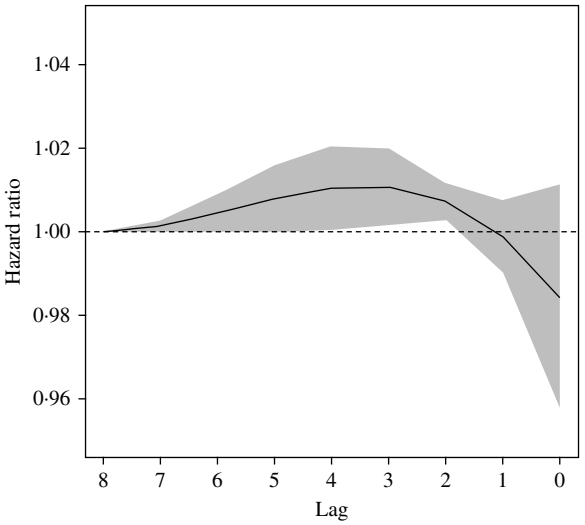


Fig. 5. Distributed lag effect: hazard ratio per °C with 95% CIs (constrained lag model at day 8), controlling for sociodemographic factors.

DISCUSSION

The average cumulative incidence of gastroenteritis over a 4-week period in those aged >5 years was over 50% higher in northern Australia in the hot and humid climatic zone (14%) compared with other climatic zones (~9%). While controlling for socioeconomic status, a statistically significant positive association was found between the probability of onset of gastroenteritis and local temperature, but no association was found between gastroenteritis and rainfall or relative humidity. There was no statistically significant interaction between temperature and climate zone or season, suggesting that the effect of temperature in the preceding week does not vary greatly by location or season. The temperature effect was highest on days 2–5 prior to the day of interest and the cumulative effect across eight previous days was 2·48% (95% CI 1·01–3·97). That is, for every 1 °C rise in temperature across the 8 days, there was a 2·48% increased risk of gastroenteritis and for a 5 °C rise across the 8 days we would expect a 13% increased risk of gastroenteritis. Given the very large total burden of gastroenteritis of over 15 million episodes per year in people aged >5 years in Australia [20] this suggests a large potential effect in this population. For average baseline days of about 41 000 cases, there could potentially be over 5000 extra cases following a 5 °C rise sustained over 8 days.

Within the outcome of total gastroenteritis episodes, there are many different viruses, bacteria and parasites that cause infectious gastroenteritis – some common pathogens in Australia include norovirus, rotavirus, *Giardia lamblia*, *Cryptosporidium*, pathogenic *E. coli*, *Campylobacter jejuni* and *Salmonella* spp. [4]. The types of pathogens are likely to vary in different parts of the world and each pathogen has its own properties, preferred habitat and possible modes of transmission to humans. Each can be expected to have its own unique association with climatic variability as temperature and other weather variables can impact differently on the replication and survival capacity of different microorganisms in different environments. For example, rotavirus is thought to be primarily spread via person-to-person contact and this virus causes a winter seasonal peak in human illness [26]. Increased temperatures are likely to have a different effect on rotavirus compared to the rates of *Salmonella* and *Campylobacter* gastroenteritis that have peak numbers in summer and spring, respectively, and are thought to be largely foodborne [6, 7, 12].

A number of studies have quantified the relationship between temperature, humidity, rainfall and specific types of diarrhoeal disease incidence across the world. For example, increased rates of *Salmonella* and *Campylobacter* infection have been consistently associated with increased temperatures [6, 7, 12, 27]. By contrast, rates of rotavirus hospitalizations in Australian children decrease with raised temperatures [8]. These examples demonstrate that it is not reasonable to extrapolate the effect of changed temperatures from one pathogen to another. However, estimating the effect of climatic variability on the outcome of all gastroenteritis has the potential to increase our understanding of the possible effect of climate change on the total burden of diarrhoea although this may well be context specific, depending on the mix of pathogens found in an area. Given the different pathogen patterns among children compared with adults it is also sensible to consider these two groups separately.

The relationship between all-cause diarrhoea and climatic variables has been examined in a few studies elsewhere using different methods and data with various strengths and weaknesses [10, 11]. The data in this study are from a survey that allowed for control of sociodemographic factors but has the potential for bias as it relied on a recall period of 4 weeks. However, a sub-study with a recall of 2 weeks suggests that the longer period was not a major cause of bias compared with the shorter period [20]. Consistent with the findings in this Australian study, a positive association was found between temperature and diarrhoea in other studies also. In a study in Peru, hospital admissions for diarrhoea in children increased by 8% for each 1 °C rise in mean temperature and there was an added El Nino effect [11]. In a Fiji study, reported diarrhoea episodes in adults were associated positively with monthly temperature and extremes of rainfall over a 20-year period [10]. In this Australian study we have used daily local postal area data which implies that interpretation of effect should be at this localized level also. The distributed lag model gave a temperature effect spread over a week which is consistent with local acquisition of the microorganism and a plausible incubation period. As the categorical variables for season and zone were not in the final model, the results should be interpreted as the effect of several days actual local temperature, not as the effect of a temperature difference over and above a variation from an average. It is possible that variations in temperature over time and place could be correlated with some other variation that occurs with seasonality or location that is not related to temperature and causes gastroenteritis. However, this seems unlikely given the combined spatio-temporal effect, the fact that temperature gave a better model fit than the variables for season and zone and the biological plausibility of temperature as a factor that could increase pathogen loads that humans are exposed to. The localized nature of the data are consistent with increased temperatures causing increases of some microorganisms in the local environs or animals, or of local temperature having an effect through inadequate local food handling or increased exposure due to changed human behaviour.

While this study demonstrates short-term effects operating at a local level, these results are still very relevant to climate change. The results suggest that hotter localized weather over a week, a scenario that is likely to be more common with climate change, is likely to increase gastroenteritis unless we adapt to counteract this. The variation by temperature while controlling for socioeconomic status suggests we have not currently 'adapted' to keep gastroenteritis risk the same across different locations and times that vary by temperature, even within a wealthy country like Australia. Much attention has been focused on malaria and vectorborne diseases; however, gastroenteritis is linked with a considerable global burden of

disease, and as shown in this study and others, is also sensitive to variations in climate. These data suggest that projections of increased temperature and changing climate conditions are likely to affect the distribution and incidence of infectious gastrointestinal disease. The association between climatic variability and gastroenteritis may well vary in different parts of the world given the probable different mix of pathogens, and different methodologies to determine associations may also lead to the need for care with interpretations. However, given this caveat, these data may be useful to assist in predicting changes in incidence in gastroenteritis with increased temperatures under climate change [28]. Those populations already at highest risk of gastroenteritis are likely to be the most affected by climate change unless adaptive strategies are undertaken.

ACKNOWLEDGEMENTS

We are grateful to OzFoodNet of the Australian Department of Health and Ageing who funded the gastroenteritis survey and the Australian Government Bureau of Meteorology for providing the climate data.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Kosek M, Bern C, Guerrant RL.** The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bulletin World Health Organization* 2003; **81**: 197–204.
2. **World Health Organization.** The World Health Report 2002 – Reducing risks, promoting healthy life. WHO Geneva, Switzerland, 2002.
3. **Mead PS, *et al.*** Food-related illness and death in the United States. *Emerging Infectious Diseases* 1999; **5**: 607–625.
4. **Hall G, *et al.*** Estimating foodborne gastroenteritis, Australia. *Emerging Infectious Diseases* 2005; **11**: 1257–1264.
5. **Scallan E, *et al.*** Prevalence of diarrhoea in the community in Australia, Canada, Ireland, and the United States. *International Journal of Epidemiology* 2005; **34**: 454–460.
6. **Kovats RS, *et al.*** Climate variability and campylobacter infection: an international study. *International Journal Biometeorology* 2005; **49**: 207–214.
7. **Kovats RS, *et al.*** The effect of temperature on food poisoning: a time-series analysis of salmonellosis in ten European countries. *Epidemiology and Infection* 2004; **132**: 443–453.
8. **D'Souza R, Hall G, Becker N.** Climatic factors associated with hospitalizations for rotavirus diarrhoea in children under 5 years of age. *Epidemiology and Infection* 2008; **136**: 56–64.
9. **Thapar N, Sanderson IR.** Diarrhoea in children: an interface between developing and developed countries. *Lancet* 2004; **363**: 641–653.
10. **Singh RB, *et al.*** The influence of climate variation and change on diarrheal disease in the Pacific Islands. *Environmental Health Perspectives* 2001; **109**: 155–159.
11. **Checkley W, *et al.*** Effect of El Nino and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. *Lancet* 2000; **355**: 442–450.
12. **D'Souza RM, *et al.*** Does ambient temperature affect foodborne disease? *Epidemiology* 2004; **15**: 86–92.
13. **Zhang Y, Bi P, Hiller J.** Climate variations and salmonellosis transmission in Adelaide, South Australia: a comparison between regression models. *International Journal of Biometeorology* 2008; **52**: 179–187.
14. **Lake IR, *et al.*** A re-evaluation of the impact of temperature and climate change on foodborne illness. *Epidemiology and Infection* 2009; **137**: 1538–1547.
15. **Haley B, Cole D, Lipp E.** Distribution, diversity, and seasonality of waterborne salmonellae in a rural watershed. *Applied and Environmental Microbiology* 2009; **75**: 1248–1255.
16. **Ashbolt R, Kirk MD.** *Salmonella* Mississippi infections in Tasmania: the role of native Australian animals and untreated drinking water. *Epidemiology and Infection* 2006; **134**: 1257–1265.
17. **OzFoodNet Working Group.** Monitoring the incidence and causes of diseases potentially transmitted by food in Australia: Annual report of the OzFoodNet Network, 2007. *Communicable Diseases Intelligence* 2008; **32**: 400–424.
18. **Australian Bureau of Statistics and Australian Institute of Health and Welfare.** The Health and Welfare of Australia's Aboriginal and Torres Strait Islander Peoples, 2005. ABS cat. 4704.0, Canberra.
19. **Hall GV, *et al.*** Frequency of infectious gastrointestinal illness in Australia, 2002: regional, seasonal and demographic variation. *Epidemiology and Infection* 2006; **134**: 111–118.
20. **Hall G, OzFoodNet Working Group.** Results from the National Gastroenteritis Survey 2001–2002. National Centre for Epidemiology & Population Health, The Australian National University Canberra, Australia, 2004.
21. **Hanigan IC, Hall G, Dear KB.** A comparison of methods for calculating population exposure estimates of daily weather for health research. *International Journal Health Geography* 2006; **5**: 38.
22. **Bureau of Meteorology.** Climate zones based on temperature and humidity. National Climate Centre, Victoria, Australia, 2005.
23. **Dear K, *et al.*** Effects of temperature and ozone on daily mortality during the August 2003 heat wave in France.

Archives Environmental Occupational Health 2005; **60**: 205–212.
24. **Schwartz J.** The distributed lag between air pollution and daily deaths. *Epidemiology* 2000; **11**: 320–326.
25. **Statacorp LP.** Intercooled Stata 9.2 for Windows. 2005. USA.
26. **Parashar U, *et al.*** Global illness and deaths caused by rotavirus disease in children. *Emerging Infectious Diseases* 2003; **9**: 565–572.
27. **Lake I, *et al.*** A re-evaluation of the impact of temperature and climate change on foodborne illness. *Epidemiology and Infection* 2009; **137**: 1538–1547.
28. **Bambrick H, *et al.*** The impacts of climate change on three health outcomes: temperature-related mortality and hospitalisations, salmonellosis and other bacterial gastroenteritis, and population at risk from dengue. Garnaut Climate Change Review, Commonwealth of Australia 2008.

5.2 Dataset 4: Meteorological data for Australian postal areas

Title: Hanigan, I. (2010). Meteorological Data for Australian Postal Areas. Australian Data Archive. doi:10.4225/13/50BBFCFE08A12

Stage of publication: Published data package at the Australian Data Archive with minted DOI

Journal Impact Factor: NA

My contribution: I was sole author of this data package. I worked alone to ensure the algorithm was correct. I conducted all data analysis and validation procedures. In accordance with the protocol I had previously developed for a paper I published in 2006 to establish this method, I assessed the sensitivity of the underlying assumptions regarding the use of population-weighted inverse distance weighted averaging. I was the corresponding author with the Australian Data Archives and the ANU Data Commons who assisted the publication of these data.

Approvals required for inclusion of Dataset 4 in this thesis:

Name	Notes	Signature/date
Mr. Ivan Hanigan		2015-09-28
Copyright	Copyright (c) 2011, The Australian National University. All rights reserved.	
Ethics Approval	Not applicable	

Australian National University

ANU Data Commons
Information Technology Services

- Welcome Guest
- Login

Meteorological Data for Australian Postal Areas

Type	Collection
Title	Meteorological Data for Australian Postal Areas
Collection Type	Dataset
Access Privileges	Meteorology and Health
DOI - Digital Object Identifier	10.4225/13/50BBFCFE08A12
Website Address	http://nesstar.ada.edu.au/webview/index.jsp?object=http://nesstar.ada.edu.au:80/obj/fStudy/au.edu.anu.ada.ddi.01230
Brief Description	Climate by areas postcode
Full Description	Background: To explain the possible effects of exposure to weather conditions on population health outcomes, weather data need to be calculated at a level in space and time that is appropriate for the health data. There are various ways of estimating exposure values from raw data collected at weather stations but the rationale for using one technique rather than another; the significance of the difference in the values obtained; and the effect these have on a research question are factors often not explicitly considered. In this study we compare different techniques for allocating weather data observations to small geographical areas and different options for weighting averages of these observations when calculating estimates of daily precipitation and temperature for Australian Postal Areas. Options that weight observations based on distance from population centroids and population size are more computationally intensive but give estimates that conceptually are more closely related to the experience of the population. Results: Options based on values derived from sites internal to postal areas, or from nearest neighbour sites; that is, using proximity polygons around weather stations intersected with postal areas; tended to include fewer stations observations in their estimates, and missing values were common. Options based on observations from stations within 50 kilometres radius of centroids and weighting of data by distance from centroids gave more complete estimates. Using the geographic centroid of the postal area gave estimates that differed slightly from the population weighted centroids and the population weighted average of sub-unit estimates. Conclusion: To calculate daily weather exposure values for analysis of health outcome data for small areas, the use of data from weather stations internal to the area only, or from neighbouring weather stations (allocated by the use of proximity polygons), is too limited. The most appropriate method conceptually is the use of weather data from sites

Status: Published
Published To:
- Australian National Data Service
Identifier: anudc:2651
Please login to request access to this dataset
Related Items

- isOutputOf: Methods for calculating population exposure estimates of daily weather [anudc:2650]
- isEnrichedBy: http://nla.gov.au/nla.party-553844
- isEnrichedBy: http://nla.gov.au/nla.party-458235
- isOwnedBy: http://nla.gov.au/nla.party-1504955
- isOwnedBy: http://nla.gov.au/nla.party-1505001
- hasAssociationWith: http://nla.gov.au/nla.party-849556
- hasAssociationWith: http://nla.gov.au/nla.party-551196
- hasDerivedCollection: New South Wales population data grouped into 15 climatic areas, 2001 [anudc:2653]
- hasDerivedCollection: Map of Bureau of Meteorology monitoring stations in NSW overlayed on Postal Area [anudc:2652]

Files

Estimates:

- Files: 1
- Size: 3.39 KB

Data Files

	within 50 kilometres radius of the area weighted to population centres, but a simpler acceptable option is to weight to the geographic centroid.
Contact Email	ivan.hanigan@anu.edu.au
Fields of Research	111706 - Epidemiology
Keywords	geographical areas health humidity population postal area postcode precipitation temperature weather Environment, Conservation, Land use postcodes rainfall zones
Date Coverage	**Date From** 1990 **Date To** 2005
Geospatial Location	**Location Type** iso31661 **Location Value** AU
Date of data creation	2010
Year of data publication	2010
Creator(s) for Citation	**Given Name** Ivan **Surname** Hanigan
Publisher for Citation	Australian Data Archive
Publications	**Identifier Type** doi **Identifier Value** 10.1017/S0950268810001901 **Publication Title** Hall G, Hanigan IC, Dear KBG and Vally H. The influence of weather on community gastroenteritis in Australia. Epidemiology and Infection. DOI: 0.1017/S0950268810001901 **Publication Reference** **Identifier Type** doi **Identifier Value** 10.1186/1476-072X-5-38 **Publication Title** Hanigan I, Hall G, Dear KBG. A comparison of methods for calculating population exposure estimates of daily weather for health research. International Journal of Health Geographics 2006;5(38) **Publication Reference** **Identifier Type**

	uri **Identifier Value** https://github.com/ivanhanigan/POAweather **Publication Title** Meteorological data for Australian Postal Areas **Publication Reference** **Identifier Type** uri **Identifier Value** http://www.garnautreview.org.au/ca25734e0016a131/WebObj/03-AThreehealthoutcomes/%24File/03-A%20Three%20health%20outcomes.pdf **Publication Title** The impacts of climate change on three health outcomes: temperature-related mortality and hospitalisations, salmonellosis and other bacterial gastroenteritis, and population at risk from dengue. H.J. Bambrick, KBG Dear, R.E. Woodruff, I C Hanigan, A.J. McMichael, Garnaut Climate Change Review, Commonwealth of Australia, (2008) **Publication Reference**
Access Rights	Available to staff and students at The Australian National University. Contact Ivan Hanigan for further information. The copyright for the material this collection is based upon is not held by The Australian National University. To utilise the items in this collection permission must be sought from the original copyright holders. Please email the contact person to gain further information about obtaining permission to utilise parts of this collection.
Rights held in and over the data	Copyright (c) 2011, The Australian National University. All rights reserved.| (c) Copyright Commonwealth of Australia 2005, Bureau of Meteorology. All rights reserved.| ABS Copyright (c) Commonwealth of Australia 2001, Australian Bureau of Statistics Census data is licensed under a Creative Commons Attribution 2.5 Australia licence. All rights reserved.

UPDATED: **13 October 2014**/RESPONSIBLE OFFICER: **Chief Information Officer** /PAGE CONTACT: **IT Services**

5.3 Paper 10: The effect on Ross River virus of proximity to waterways

Title: Vally, H., Peel, M., Dowse, G. K., Cameron, S., Codde, J. P., Hanigan, I., and Lindsay, M. D. A. (2012). Geographic Information Systems used to describe the link between the risk of Ross River virus infection and proximity to the Leschenault Estuary, WA. Australian and New Zealand Journal of Public Health, 36(3), 229235. doi:10.1111/j.1753-6405.2012.00869.x

Stage of publication: Published peer-reviewed paper

Journal Impact Factor: 2

My contribution: In this paper I worked with lead author Dr Vally to select the statistical approach appropriate to the research question and data. I considered a simulation approach before selecting the use of Poisson regression on spatial subsets of data as the method to use. I implemented this modelling framework and adjusted for the urban or rural context by geographical region. The Geographical Information System (GIS) data preparation was performed previously by the other researchers and I needed to recreate some of the analytical data from the original data sources, without access to some of the previously derived analysis data. I created all the maps and graphs that were used in the paper. I responded to the statistical questions from the reviewers who wanted justification for the use of the modelling framework, which I was able to provide.

Approvals required for inclusion of Paper 10 in this thesis:

Name	Notes	Signature/date
Dr. Hassan Vally		2015-08-19
Mr. Mark Peel		Could not be contacted 2015-12-05
Dr. Gary K. Dowse		2015-09-30
Prof. Scott Cameron		2015-10-01
Prof. Jim P. Codde		2015-11-05
Mr. Ivan Hanigan		2015-09-28
Dr. Michael D. A. Lindsay		2015-09-29
Copyright	License Number: 3723610619805 License date: Oct 07, 2015 Licensed Content Publisher: John Wiley and Sons	2015-10-07
Ethics Approval	No ethics approval was required as this work resulted from an analysis of routine notification data at the Western Australian Department of Health.	

Article

Methods

Geographic Information Systems used to describe the link between the risk of Ross River virus infection and proximity to the Leschenault estuary, WA

Hassan Vally

Communicable Disease Control Directorate, Department of Health, Western Australia; National Centre for Epidemiology and Population Health, Australian National University, ACT; School of Public Health & Human Biosciences, La Trobe University, Victoria

Mark Peel

South Metropolitan Area Health Service, Department of Health, Western Australia

Gary K. Dowse

Communicable Disease Control Directorate, Department of Health, Western Australia

Scott Cameron

National Centre for Epidemiology and Population Health, ANU College of Medicine and Health Sciences, Australian National University, ACT

Jim P. Codde

South Metropolitan Area Health Service, Department of Health, Western Australia; Curtin Health Innovation Research Institute, Faculty of Health Sciences, Curtin University of Technology, Western Australia

Ivan Hanigan

National Centre for Epidemiology and Population Health, ANU College of Medicine and Health Sciences, Australian National University, ACT

Michael D.A. Lindsay

Environmental Health Hazards Unit, Department of Health, Western Australia

Abstract

***Objective*:** To investigate the relationship between risk of Ross River virus (RRV) infection and proximity to mosquito-breeding habitat surrounding a tidal wetland ecosystem in south-west Australia.

Methods: Geographic information systems (GIS) were used to spatially map cases of RRV disease in the Leschenault region between July 1995 and June 1996. Half kilometre buffer zones were constructed around the Leschenault Estuary and associated waterways; RRV disease case counts were calculated for each zone.

Results: Different relationships between RRV disease incidence and proximity to saltmarsh mosquito habitat were observed east of the Leschenault Estuary compared with an urban region to the south. Disease incidence showed a decreasing trend away from eastern margins of the Estuary, particularly for the first 2 km. In the urban region, RRV disease risk was low close to the Estuary, but increased further out and remained steady across the remainder of that region.

Conclusions: The findings support an increased risk of contracting RRV disease for people residing close to eastern margins of the Leschenault Estuary.

Implications: This study highlights how historical data combined with GIS can improve understanding of the epidemiology of RRV disease. This has a valuable role in assessing the risk of mosquito-borne disease for land-use planning.

Key words: Ross River virus, proximity to waterway, Geographic Information Systems, risk assessment, Western Australia

Aust NZ J Public Health. 2012; 36:229-35

doi: 10.1111/j.1753-6405.2012.00869.x

Ross River virus (RRV) is a mosquito-borne alphavirus that causes outbreaks of polyarthritic disease in humans in most regions of Australia.[1] It is generally most active in the south-west of Western Australia (WA) between September and May.[2,3] The extent of this activity, however, is largely dependent on environmental conditions that favour vector mosquito breeding and other key components of the transmission cycle.[4-6] Previous epidemics of RRV in the south-west of WA have been preceded by high tides or unusually heavy late spring and summer rains, which have enabled the breeding of large numbers of *Aedes camptorhynchus* (Thomson) and/or *Aedes vigilax* (Skuse) mosquitoes in seasonal wetlands and tidal salt marshes along the south-west coast of WA.[6,7]

Despite the fact that mosquitoes are known to breed in these ecosystems, residential developments continue to be built close to these wetlands where problems with RRV disease would be expected to occur.

Submitted: July 2011 **Revision requested**: November 2011 **Accepted**: January 2012

Correspondence to: Dr Michael Lindsay, Environmental Health Directorate, Department of Health, PO Box 8172, Perth Business Centre, WA, 6849; e-mail: mike.lindsay@health.wa.gov.au

© 2012 The Authors. ANZJPH © 2012 Public Health Association of Australia

This highlights the need to better understand and document the relationship between RRV disease incidence and proximity to wetlands. Such information will be an important resource for local governments and planning authorities to manage health risks for new residential developments near natural mosquito breeding habitats and to guide implementation of mosquito management programs. The work will also contribute to our understanding of how areas at risk of vector-borne disease may change if predicted climate-associated changes in the boundaries of tidal wetlands[8] occur.

Geographic Information Systems (GIS) allow the input, storage, manipulation and analyses of spatial data.[9,10] GIS have a number of public health applications and are especially useful in examining relationships between disease incidence and environmental variables.[10] GIS have been used to investigate spatial relationships of a number of vector-borne diseases, their vectors and environmental factors predisposing to outbreaks.[11-15] Indeed, Muhar et al.[15] used GIS to identify relationships between the proportion of wetlands and bushland in suburbs of Brisbane and the infection rate of RRV, while Ryan et al.[16] demonstrated a spatial relationship between incidence of RRV disease and intense mosquito biting activity in Redland Shire, southern Queensland. However, the relationship between the incidence of RRV disease and distance from wetland ecosystems has not, to our knowledge, been examined in detail.

This present study, therefore, aimed to develop a methodological approach for examining the spatial relationship between the incidence of RRV disease and proximity to a mosquito-breeding habitat – the Leschenault Estuary, a large wetland ecosystem north of Bunbury, WA. We took advantage of some of the most complete RRV disease data available, from one of the largest outbreaks of RRV in WA that occurred in the south-west in 1995-1996.[7] It represented a powerful notification dataset, including enhanced surveillance to determine likely place of exposure to RRV-infected mosquitoes. Data were collected in a form amenable to precise geocoding. In addition, given that the boundaries of the Leschenault Estuary have not changed between this data being collected and analysed, these findings remain relevant to understanding the local epidemiology of RRV disease around this estuary.

Methods

Study area

Two major regions in south-west Western Australia were affected by an outbreak of RRV disease in 1995-96: Leschenault and Capel-Busselton, on the Swan Coastal Plain between 165-190 km and 190-230 km, respectively, south of Perth. Detailed descriptions of the geography of both regions and aspects pertaining to mosquito breeding and ecology of vector-borne diseases have been described previously.[17]

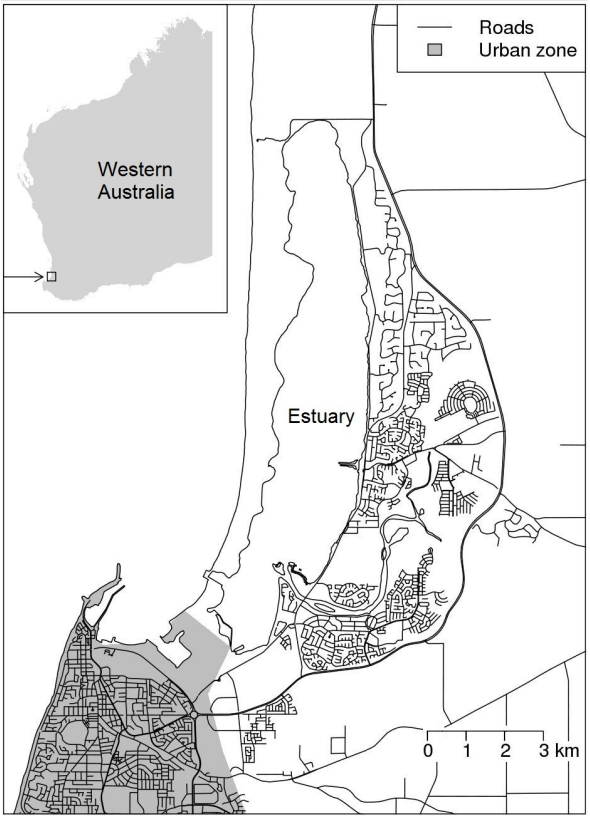
The Leschenault Estuary was chosen as the main focus for the study, as a considerable proportion of the RRV disease cases in this epidemic were close to this waterway and the estuary had a well-defined boundary. The estuary is one of the largest waterways in south-west WA. Two river systems (Collie and Brunswick Rivers) flow into the estuary and a third (Preston River) has been diverted into it (Figure 1). The estuary is connected to the Indian Ocean by a man-made cut and, along with the lower reaches of the rivers that flow into it, is subject to tidal influences. The margins of these waterways are characterised by substantial areas of tidal saltmarsh. The estuary and associated waterways are used extensively for recreational activities including fishing, swimming and boating.

Disease data

A dataset of human cases of RRV disease from the Leschenault, Capel and Busselton regions in south-west WA between 1 July 1995 and 30 June 1996 was compiled from laboratory-confirmed cases. The criteria for notification of RRV disease cases to the Department of Health (DoH) required detection of virus in a clinical specimen, a fourfold or greater rise in serum antibody titres between acute and convalescent sera, or the detection of IgM specific antibodies. Cases were notified to the DoH by medical practitioners and/or diagnosed by the WA Centre for Pathology and Medical Research (PathCentre WA). Each case was recorded by date of onset rather than date of notification in order to provide a more accurate estimate of the timing of exposure to infected mosquitoes. However, for cases in which an estimate of the date of onset was not recorded, this was assumed to be the same as the date of notification.

A total of 524 cases were included in the dataset defined above.

Figure 1: Map showing the Leschenault Estuary and surrounds.



© 2012 The Authors. ANZJPH © 2012 Public Health Association of Australia

Of these, 389 cases (74.2 %) were interviewed by local government environmental health officers using an enhanced surveillance questionnaire to facilitate collection of additional data indicating potential places and timing of exposure to RRV. These data, in turn, were interpreted by experienced mosquito-borne disease control staff to determine a most likely time and location of exposure, taking into account incubation period of RRV, recall of exposure to biting mosquitoes, travel history and relative risk of RRV disease in different regions of Western Australia at the time. Cases were included in this dataset if the Leschenault, Capel or Busselton areas were indicated as the likely place of exposure to RRV. For those cases for which information about likely place of exposure was not available, cases were included in the dataset if the place of residence was within the study region.

Geocoding of disease data

The dataset created for spatial analyses consisted of all cases for which data relating to place of exposure or residential address could be precisely geocoded. That is, if exposure data were collected and geocodable to a specific cadastral lot (a legally defined property boundary), the co-ordinates of the cadastral centroid were ascribed to the case. In instances where residential data was given precisely (street number and name) and exposure data was indicated as the same street (but with no street number), it was assumed that these locations were identical, and the co-ordinates of the residential cadastral lot centroid were ascribed to the case. If exposure data could not be precisely geocoded (or was not available) and residential address was geocodable to an exact location (and within the region of interest) then the co-ordinates for the residential cadastral lot centroid were ascribed to the case. In addition, if place of exposure data or residential data were not given as an exact location but could be pinpointed with reasonable confidence (e.g. a street corner within 250 metres) then these cases were also geocoded. All other cases were excluded from the dataset for spatial analyses. Thus, the size of the dataset for spatial analyses was 371 cases.

Waterway analyses

To investigate the relationship between proximity to waterways where mosquito-breeding habitat was known to occur and the incidence of RRV disease, spatially-mapped waterways from orthorectified aerial imagery were acquired from the Department of Land Information (Government of Western Australian) for sections within the study area. This imagery was registered and imported into the MapInfo® GIS software (MapInfo Professional Version 7.0). The waterways of interest, as identified by Wright,[18] were digitised from the registered imagery into polygons for buffering analyses.

Because of the influence of tidal inundation of salt marshes along the lower reaches of the river systems connected to the Leschenault Estuary, these systems were included in waterway analyses. That is, the Leschenault Estuary was extended to include short sections of the Collie and Brunswick River systems as well as a section of the Preston River subject to tidal flooding. Half-kilometre buffer zones were constructed up to a distance of 10 km from the high water mark of the Leschenault Estuary and associated waterways, and RRV disease case counts were calculated for each of these zones.

The estimated population for each of these buffer zones was based on the 1996 ABS Census Data. The collection district (CD) boundaries were split to the buffer zones based on geographic overlap. The split CDs were then aggregated to estimate the total population per 500-metre buffer from the Estuary. Incidence rates for each buffer zone were calculated by dividing the number of cases within the buffer zone by the estimated residential population within this zone. In order to better delineate the relationship between RRV disease incidence and proximity to the Leschenault Estuary, separate analyses were conducted for urban Bunbury and the area east of the Leschenault Estuary.

RRV incidence and population data were initially determined in buffers out to 10 km from the estuary. However, the analysis was limited to 7.5 km in view of the fact that beyond 7.5 km east of the estuary the main land-use is broad acre farming and the resident population was too low to enable calculation of statistically robust attack rates. Additionally, a different ecosystem with sources of mosquitoes from other wetland habitats occurs beyond 7.5 km south of Bunbury. Therefore including data beyond that distance would have introduced confounding factors.

The association between RRV infection and distance from the waterway was examined using a Poisson generalised linear model. We fitted a log-linear model to counts of infected individuals with buffer zone as the predictor. Population was included as an offset. We estimated the model and point-wise 95% confidence intervals using the statistical software package R version 2.10.1 (R Development Core Team, Vienna, Austria, 2009).

Results

Descriptive epidemiology

The major outbreak of RRV disease that occurred in the south-west of WA during 1995/96 has been described previously.[7] Within the Leschenault and Capel-Busselton regions there were 524 cases of RRV disease notified between July 1995 and June 1996, with an attack rate estimated to be 0.74%. A total of 273 males and 251 females were clinically infected during this outbreak, yielding a ratio of male to female cases of 1.1:1.0. Information regarding the age of infected individuals was available for 517 cases, with RRV disease found to most commonly affect those between 40 and 49 years of age. The outbreak peaked in January and February 1996 (244 and 136 cases, respectively).

Criteria for ascribing co-ordinates

The number (and percentage) of cases in each category used to ascribe co-ordinates for the 371 cases in the spatial analysis were as follows: place of exposure geocoded to a cadastre lot: 48 (12.9%); residential address precisely geocodable and the same as place of exposure: 64 (17.3%); residential address geocoded to a cadastre lot: 247 (66.6%); place of exposure or residential address geocoded to street level (within 250 metres): 12 (3.2%). Thus, only 12.9% of

© 2012 The Authors. ANZJPH © 2012 Public Health Association of Australia

cases included in the analysis were known to have resided outside the study region and at least 83% of geocoded cases resided within the study region. Therefore, although subsequent calculations of infection rates within 500 m buffers around waterways were based on resident population only, the potential error introduced by including non-resident cases (who were exposed in the region) is likely to have been small. The 25.8% of RRV cases for which no exposure data was available were evenly distributed throughout the study region so it is unlikely that this was a further source of uncertainty in calculating exposure rates.

Waterway analyses

Table 1 summarises the crude RRV disease incidence rate per 1,000 persons calculated for each buffer zone up to 7.5 km from the Leschenault Estuary. Disease incidence showed no obvious pattern when all cases surrounding the Estuary were considered. However, more detailed inspection of the spatial plot of RRV disease cases suggested distinct clusters of high incidence within the area surrounding the Leschenault Estuary (Figure 2). Specifically, the cases to the east of the Estuary appeared to cluster separately to those in urban Bunbury. Therefore, to further delineate the relationship between proximity to waterway and RRV disease incidence, separate analyses for the area east of the Leschenault Estuary and urban Bunbury were undertaken. Aggregating the smaller urban residential lots of Bunbury separately from larger rural lots created the region for urban Bunbury. The eastern extent of the Leschenault Estuary included the entire region to the east and north of the estuary, with the southern boundary of this subregion bordering on the urban lot grouping of the Bunbury subregion (Figure 2).

Figures 3 and 4 show the empirical relationship between proximity to the Leschenault Estuary and RRV disease incidence for the area east of the Estuary and for urban Bunbury, respectively, as well as Poisson regression models. The removal of the influence of urban Bunbury suggested a strong RRV disease incidence gradient for the first 2 km from the Leschenault Estuary for those cases east of the estuary (Figure 3). Beyond this, despite an aberration at 2.5 km from the estuary where the incidence rate spiked to 8.5 cases per 1,000, both the number of cases and resident populations in buffers fell considerably, resulting in much lower incidence rates and less robust estimates of incidences. Poisson regression confirmed a disease incidence gradient for those cases east of the estuary, with a 22% reduction in the rate of disease per km increased distance from the Estuary (p=0.002).

In contrast, the examination of the empirical relationship between proximity to the Leschenault Estuary and RRV disease incidence for urban Bunbury provided little evidence for the influence of the estuary on risk of disease (Figure 4). Indeed, the results suggest that the risk of disease was actually lower within 1.5 km of the southern end of the Estuary and that thereafter the risk of disease increased but showed no discernable pattern across the urban area. Poisson regression was consistent with this, with no significant relationship found (p=0.544).

Discussion

The major objective of this investigation was to develop methodology to investigate the relationship between close proximity to particular wetlands and an increased risk for RRV disease. Although such a relationship has a solid ecological basis, there are few studies that have addressed this issue in detail. Credible scientific evidence demonstrating this association is required in order to inform local government and planning authorities of high risk wetlands so that appropriate strategies to minimise the risk of contracting RRV disease in areas close to these wetlands can be implemented. Additionally, more detailed information regarding the nature of the relationship between the proximity to wetlands

Table 1: RRV cases, no of persons and RRV disease incidence rate per 1000 persons for each 500m buffer in the entire (combined) Leschenault region and eastern Estuary and Urban Bunbury subregions.

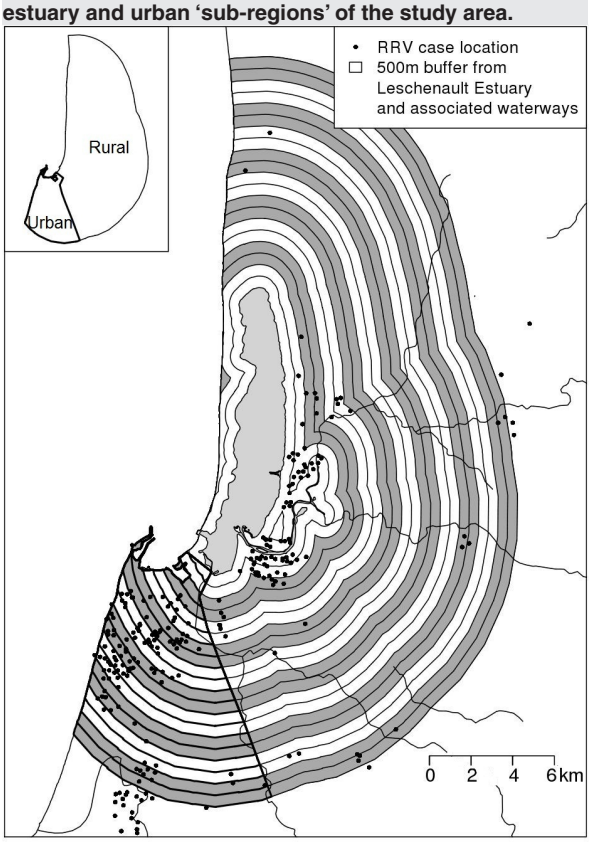
	Entire Leschenault region			Eastern estuary region			Urban Bunbury region		
Buffer (km)	RRV cases	Total persons	RRV rate/1000	RRV cases	Total persons	RRV rate/1000	RRV cases	Total persons	RRV rate/1000
0.5	46	6,048	7.6	46	6,048	7.6	0	0	-
1.0	23	4,023	5.7	23	4,004	5.7	0	19	0.0
1.5	10	1,867	5.4	10	1,858	5.4	0	9	0.0
2.0	6	1,512	4.0	3	967	3.1	3	545	5.5
2.5	11	1,988	5.5	6	705	8.5	5	1,283	3.9
3.0	6	2,220	2.7	2	571	3.5	4	1,649	2.4
3.5	7	2,485	2.8	1	413	2.4	6	2,072	2.9
4.0	18	2,899	6.2	0	408	0.0	18	2,491	7.2
4.5	15	2,559	5.9	1	343	2.9	14	2,216	6.3
5.0	9	2,835	3.2	0	402	0.0	9	2,433	3.7
5.5	15	3,273	4.6	1	538	1.9	14	2,735	5.1
6.0	9	2,406	3.7	0	398	0.0	9	2,008	4.5
6.5	14	2,192	6.4	0	337	0.0	14	1,855	7.5
7.0	10	2,130	4.7	0	188	0.0	10	1,942	5.1
7.5	7	1,862	3.8	3	393	7.6	4	1,469	2.7
Totals	**206**	**40,299**	**5.1**	**96**	**17,573**	**5.5**	**110**	**22,726**	**4.8**

© 2012 The Authors. ANZJPH © 2012 Public Health Association of Australia

and RRV infection risk is required in order to focus prevention and control strategies. Finally, the associated changes in incidence and distribution of vector-borne diseases that may occur in Australia as a result of changing climate[8] is a further driver to better understanding the epidemiology of these diseases.

The epidemic of RRV disease in south-west Western Australia in 1995-1996[7] is one of the largest epidemics on record in that region and provided an ideal source of high quality surveillance data for this project. Despite the fact that this data is some 15 years old, it provided a unique dataset containing geocodable notification data in order to explore the use of GIS to examine the epidemiology of RRV disease surrounding wetlands known to be mosquito-breeding habitats. Furthermore, the Leschenault Estuary appeared to provide an ideal wetland around which to examine RRV infection risk, as there were a large number of cases in the area surrounding this wetland during this outbreak and this mosquito breeding habitat had been well defined.[18] As GIS technology facilitates the easy measurement of distances between multiple features of interest, this provided the ideal tool to measure the distances between the locations of multiple RRV disease cases and the irregular perimeter of the Leschenault Estuary. Thus, incidence rates for RRV disease and proximity to this wetland could be examined in a straightforward manner.

Figure 2: Map of Leschenault region showing location of geocoded RRV disease cases, 0.5 km buffers around major tidal waterways and boundaries of the eastern estuary and urban 'sub-regions' of the study area.



The major finding from this study was evidence that the risk of RRV disease in the Leschenault region is elevated in close proximity to the Leschenault Estuary. Analyses focusing on the region east of the Leschenault Estuary suggested a strong disease risk gradient for the first 2 km from the estuary. Poisson regression confirmed a relationship between disease incidence and distance from the

Figure 3: Incidence of RRV disease per 1,000 persons in 500m buffers away from the Leschenault Estuary and tidal sections of the Collie, Brunswick and Preston Rivers, for the eastern Estuary subregion. Poisson regression model with 95% confidence intervals is also shown.

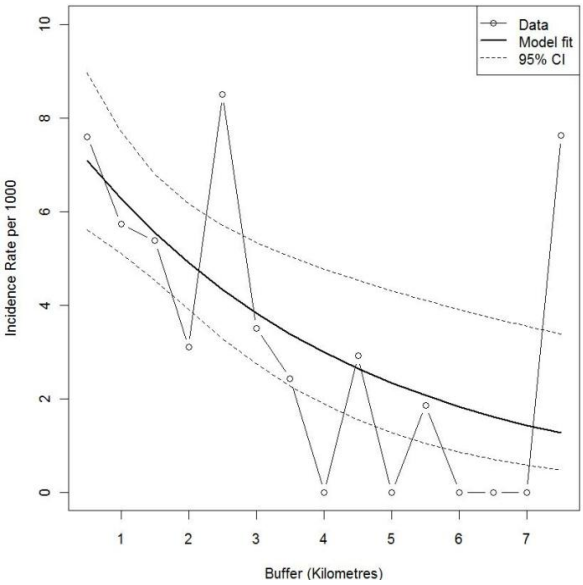
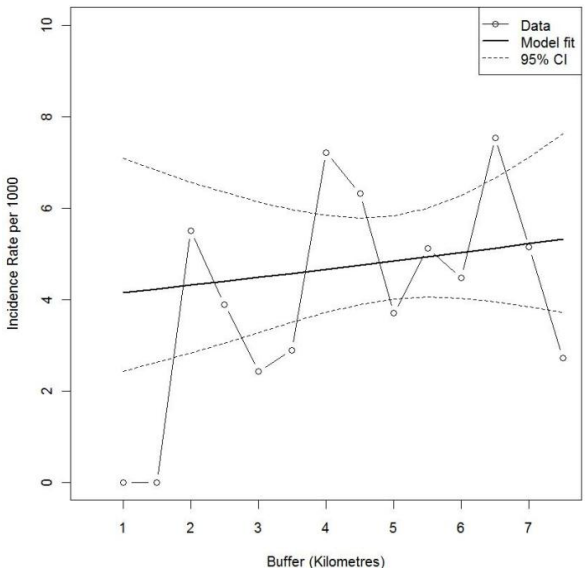


Figure 4: Incidence of RRV disease per 1,000 persons in 500m buffers away from the Leschenault Estuary and tidal sections of the Collie, Brunswick and Preston Rivers, for the urban Bunbury subregion. Poisson regression model with 95% confidence intervals is also shown.



© 2012 The Authors. ANZJPH © 2012 Public Health Association of Australia

estuary for this region. The presence of an elevated risk of disease close to the estuary which decreased as one moved further away was consistent with our expectations and our understanding of the ecology of RRV disease in this region in which the saltmarsh mosquito *Ae. camptorhynchus* is a major vector.[7]

In contrast, analyses of the urban Bunbury region did not provide evidence supporting an influence of the estuary on disease risk, with the disease risk being very low closest to the southern end of the estuary and then increasing and remaining relatively constant across the remainder of the urban area. The apparent lack of influence of the estuary on RRV disease in Bunbury may reflect the fact that mosquito-breeding habitat along the southern margin of the estuary has all but disappeared as a result of construction of Bunbury Harbour and associated infrastructure.

Two explanations for the observation of a relatively even risk of RRV disease across the remainder of the urban environment seem most likely. Firstly, this may reflect the influence of urban or peridomestic mosquitoes within urban Bunbury. *Aedes notoscriptus* (Skuse), a known vector species[19] with limited dispersal capability[20] has been collected in Bunbury during outbreaks of RRV disease and has yielded RRV (Lindsay, M.D.A., unpublished results). Alternatively, one or more fresh water wetlands within urban Bunbury, not included in these analyses, may have enabled breeding of secondary vectors of RRV, such as *Culex annulirostris* (Skuse),[19] which subsequently dispersed into surrounding suburbs during the outbreak. Spatial distribution of both species was found to differ from that of saltmarsh mosquitoes in previous studies in urban environments in eastern Australia.[21]

In order to perform the analyses described in this study, buffer zones were created around the Leschenault Estuary. RRV cases within these buffer zones were then counted and the resident population estimated in order to calculate incidence rates. Although the Leschenault Estuary appeared to be a good choice for the waterway analyses, based on the availability of spatial data accurately describing the perimeter of the waterway, and the relatively large number of RRV disease cases surrounding it, the choice of this estuary for these analyses ultimately proved to be problematic. Incidence rates for RRV disease in many of the buffer zones surrounding this waterway were actually low, with large associated uncertainties, due to the large size of the estuary and hence the large area encompassed by buffer zones and the relatively small number of residents within several buffer zones surrounding the estuary. In addition, the area surrounding the Leschenault Estuary represented a complex ecosystem. Urban Bunbury was located to the south and clearly represented a separate and distinct environment when compared with the area east of the Leschenault Estuary. Compounding this, several river systems fed into the estuary and therefore a number of arbitrary judgements needed to be made as to how to best define the perimeter of the waterway for the purpose of analysis, which clearly added an additional source of uncertainty to the analyses. Finally, there were several small water bodies within the area encompassed by the buffer zones surrounding the Leschenault Estuary, which were not taken into account in the analyses, and these may have confounded the examination of the relationship between RRV disease risk and proximity to the Leschenault Estuary.

In addition to limitations that were directly linked to the choice of the Leschenault Estuary, there were also limitations inherent in the study methodology. First, the calculation of the populations residing within buffer zones was subject to a considerable amount of uncertainty. Populations within buffer zones were estimated by the extent of geographic overlap between census collection districts and these zones. This clearly represented a crude approximation and therefore may have been a significant source of error when calculating incidence rates. Second, the study region, being a tourist area, was subject to the large influx of people on holiday (particularly during summer, when most cases were recorded), which could not be factored into the calculation of incidence rates within buffer zones.

It is important to note that data for this project were obtained from an enhanced surveillance database which contained more detailed information for those cases for which enhanced surveillance interviews were completed. This included a more accurate assessment of the date of onset of RRV disease as well as the recording of the place at which cases were most likely to have been infected with RRV. The collection of this enhanced surveillance data is important in the context of RRV disease, as individuals were frequently shown to have been infected at locations other than their place of residence, whilst visiting areas close to mosquito breeding sites, and thus routine notification data may provide a false representation of the epidemiology of the disease, which then may misinform risk assessment, prevention and control measures.

Despite the limitations, the present study provides a useful starting point for refining methodologies for examining the relationship between proximity to waterways and risk of RRV disease infection. Specifically, if surveillance data is to be used for spatial analyses, then a greater effort needs to be made to collect place of exposure data (or residential data) in a form that is geocodable. It is also important that there are sufficient disease cases in the dataset for the period being analysed (so that there is sufficient power to delineate any spatial relationships that may exist). Additionally, sufficiently detailed information about ecologically relevant perimeters of the waterways of interest is also important, so that the boundaries which define the waterway for the purpose of analyses can be determined with confidence. Furthermore, a good understanding of the local ecology of the land surrounding the waterway is required in order to direct analyses and to assist in the interpretation of findings.[22]

Finally, it may need to be accepted that there will always be uncertainties associated with the determination of the underlying population within buffer zones for the calculation of incidence rates. Strategies to minimise these uncertainties need to be explored for future studies. For example, GIS can quickly calculate the number of dwellings within a given area. Therefore, it may be simpler and more accurate to calculate incidence rates per dwelling within these 'artificial' boundaries for similar studies in the future. However, if the region of interest is a tourist area, such as was the case in the present study, then the transient influx of people into the study

© 2012 The Authors. ANZJPH © 2012 Public Health Association of Australia

area will also present a challenge in the accurate calculation of incidence rates.

In conclusion, the present study provided evidence suggesting that there was an increased risk of contracting RRV disease associated with either residing, or pursuing activities, in close proximity to Leschenault Estuary. Despite the fact that salt marsh mosquitoes are capable of travelling several kilometres,[23] the results of the analysis suggested that in this instance the increased risk of contracting RRV disease was strongly associated with a proximity of 1-2 km from the eastern boundary of the Estuary. Clearly, other factors such as vegetation (providing harbourage and dispersal corridors for vector mosquitoes) and prevailing winds, as well as the mosquito species themselves, can influence the distances that mosquitoes may travel. Thus future studies should consider these factors, together with the distribution, abundance and immune status of vertebrate hosts.

Acknowledgements

The authors thank colleagues in the Communicable Diseases Control and Environmental Health Directorates as well as Stephen Spiker from the Epidemiology Branch of the Western Australian Department of Health for their assistance with this study. Thanks also to Niels Becker and Mark Clements from the National Centre of Epidemiology and Population Health at The Australian National University for statistical advice. We acknowledge the Environmental Health Officers from Western Australian Local Governments who collected enhanced surveillance data and those individuals who contracted Ross River virus disease and subsequently provided data for this study. This work was completed as part of the requirements for the Masters of Applied Epidemiology, which was funded by the Commonwealth Department of Health and Ageing.

References

1. Harley D, Sleigh A, Ritchie S. Ross River virus transmission, infection, and disease: a cross-disciplinary review. *Clin Microbiol Rev*. 2001;14:909-32.
2. Lindsay MDA. Ross River and Barmah Forest viruses. Mosquito Management Manual: Darwin (AUST): Western Australian Department of Health; 2011. p. 200-28.
3. Lindsay MDA, Breeze AL, Harrington SA, Johansen CA, Broom AK, Gordon C, et al. Ross River and Barmah Forest viruses in Western Australia, 2000/01 – 2003/04: contrasting patterns of disease activity. *Arbo Res Aust*. 2005;9:194-201.
4. Mackenzie JS, Lindsay MD, Daniels P. The effect of climate on the incidence of vector-borne viral diseases in Australia: the potential value of seasonal forecasting. In: Hammer G, Nicholls N, Mitchell C, editors. *Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems – The Australian Experience*. Dordrecht (NLD): Kluwer Academic Publishers; 2000. p. 429-52.
5. Lindsay MD, Mackenzie JS, Condon RJ. Ross River virus outbreaks in Western Australia: Epidemiological aspects and the role of environmental factors. In Ewan CE, Bryant EA, Calvert GD, Garrick JA, editors. *Health in the Greenhouse*. Canberra (AUST): AGPS; 1993. p. 85-100.
6. Woodruff RE, Guest CS, Garner MG, Becker N, Lindsay MDA. Early warning of Ross River virus epidemics: combining climate and mosquito surveillance data. *Epidemiology*. 2006;17(5):1-7.
7. Lindsay MD, Olivera N, Jasinska E, Johansen C, Harrington S, Wright T, et al. Western Australia's largest recorded outbreak of Ross River virus disease. *Arbo Res Aust*. 1997;7:147-52.
8. Russell RC. Mosquito-borne disease and climate change in Australia: time for a reality check. *Aust J Entomol*. 2009;48:1-7.
9. Glass GE. Geographic Information Systems. In: Nelson KE, Williams CM, Graham NMH, editors. *Infectious Disease Epidemiology*. Gaithersburg (MD): Aspen Publishers; 2001. p. 231-53.
10. Ricketts TC. Geographic information systems and public health. *Annu Rev Public Health*. 2003;24:1-6.
11. Allen TR, Wong DW. Exploring GIS, spatial statistics and remote sensing for risk assessment of vector-borne diseases: A West Nile virus example. *Int J Risk Assess Manag*. 2006;6:253-75.
12. Eisen L, Eisen R. Need for improved methods to collect and present spatial epidemiologic data for vector-borne diseases. *Emerg Infect Dis*. 2008;13:1336-7.
13. Mushinzimana E, Munga S, Minakawa N, Li L, Feng C-C, Bian L, et al. Landscape determinants and remote sensing of Anopheline mosquito larval habitats in the western Kenya highlands. *Malar J*. 2006;5:1-11.
14. Whelan P, Merianos A, Hayes G, Krause V. Ross River virus transmission in Darwin, Northern Territory, Australia. *Arbo Res Aust*. 1997;7:337-45.
15. Muhar A, Dale PER, Thalib L, Arito E. The spatial distribution of Ross River virus infections in Brisbane: Significance of residential location and relationships with vegetation types. *Environ Health Prev Med*. 2000;4:184-9.
16. Ryan PA, Alsemgeest D, Gatton ML, Kay BH. Ross River Virus Disease Clusters and Spatial Relationship with Mosquito Biting Exposure in Redland Shire, Southern Queensland, Australia. *J Med Entomol*. 2006;43:1042-59.
17. Lindsay MDA. *Ecology and Epidemiology of Ross River Virus in Western Australia* [PhD]. Perth (AUST): University of Western Australia; 1995.
18. Wright AE. *Report on the Mosquito Eradication Campaign: Survey of Mosquitoes in the Bunbury Region, WA*. Perth (AUST): Western Australian Department of Health; 1986.
19. Russell RC. Ross River virus: ecology and distribution. *Annu Rev Entomol*. 2002;47:1-31.
20. Watson TM, Saul A, Kay BH. *Aedes notoscriptus* (Diptera: *Culicidae*) survival and dispersal estimated by mark-release-recapture in Brisbane, Queensland, Australia. *J Med Entomol*. 2000;37:380-4.
21. Ryan PA, Lyons SA, Alsemgeest D, Thomas P, Kay BH. Spatial statistical analysis of adult mosquito (Diptera: *Culicidae*) counts: an example using light trap data in Redland Shire, Southeastern Queensland, Australia. *J Med Entomol*. 2004;41:1143-56.
22. Kelly-Hope LA, Purdie DM, Kay BH. Ross River Virus Disease in Australia, 1886–1998, with Analysis of Risk Factors Associated with Outbreaks. *J Med Entomol*. 2004;41:133-50.
23. Liehne PFS. *An Atlas of the Mosquitoes of Western Australia*. Perth (AUST): Western Australian Department of Health; 1991.

© 2012 The Authors. ANZJPH © 2012 Public Health Association of Australia

5.4 Software package 3: Reproducible code for paper 10

Title: Hanigan, I.C. (2015). Reproducible code for Vally et al. 2012 Ross River Rates. https://github.com/ivanhanigan/Vally2012-RossRiverRates

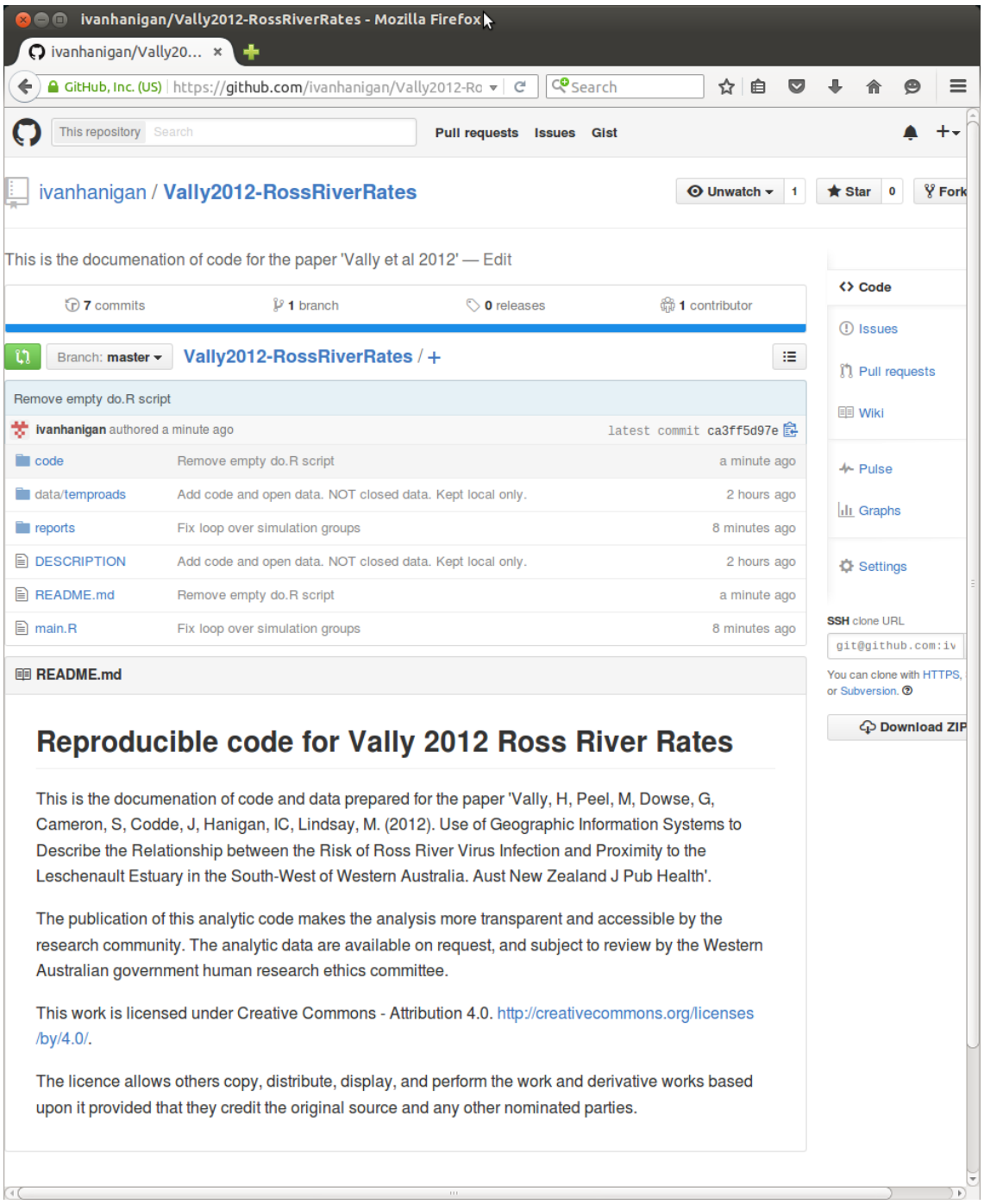
Stage of publication: Open source software available on Github.

Journal Impact Factor: NA

My contribution: I was the sole author on this software package that provides reproducible code for paper 10 of this thesis (Vally et al 2012) . The publication of this code makes the analysis more transparent and accessible by the research community. The analytic data are available on request, and subject to review by the Western Australian Government human research ethics committee.

Approvals required for inclusion of Software package 3 in this thesis:

Name	Notes	Signature/date
Mr. Ivan Hanigan		2015-09-28
Copyright	Creative Commons - Attribution 4.0. The licence allows others to copy, distribute, display, and perform the work and derivative works based upon it provided that they credit the original source.	
Ethics Approval	Not applicable	


ivanhanigan/Vally2012-RossRiverRates - Mozilla Firefox
ivanhanigan/Vally20...
GitHub, Inc. (US) https://github.com/ivanhanigan/Vally2012-Ro
Search
This repository Search
Pull requests Issues Gist
ivanhanigan / Vally2012-RossRiverRates
Unwatch 1 Star 0 Fork
This is the documenation of code for the paper 'Vally et al 2012' — Edit
7 commits 1 branch 0 releases 1 contributor
Branch: master Vally2012-RossRiverRates / +
Remove empty do.R script
ivanhanigan authored a minute ago latest commit ca3ff5d97e
code Remove empty do.R script a minute ago
data/temproads Add code and open data. NOT closed data. Kept local only. 2 hours ago
reports Fix loop over simulation groups 8 minutes ago
DESCRIPTION Add code and open data. NOT closed data. Kept local only. 2 hours ago
README.md Remove empty do.R script a minute ago
main.R Fix loop over simulation groups 8 minutes ago
Code
Issues
Pull requests
Wiki
Pulse
Graphs
Settings
SSH clone URL
git@github.com:iv
You can clone with HTTPS, or Subversion.
Download ZIP
README.md
Reproducible code for Vally 2012 Ross River Rates
This is the documenation of code and data prepared for the paper 'Vally, H, Peel, M, Dowse, G, Cameron, S, Codde, J, Hanigan, IC, Lindsay, M. (2012). Use of Geographic Information Systems to Describe the Relationship between the Risk of Ross River Virus Infection and Proximity to the Leschenault Estuary in the South-West of Western Australia. Aust New Zealand J Pub Health'.
The publication of this analytic code makes the analysis more transparent and accessible by the research community. The analytic data are available on request, and subject to review by the Western Australian government human research ethics committee.
This work is licensed under Creative Commons - Attribution 4.0. http://creativecommons.org/licenses/by/4.0/.
The licence allows others copy, distribute, display, and perform the work and derivative works based upon it provided that they credit the original source and any other nominated parties.

Chapter 6

Discussion and conclusions

The aims of this PhD were twofold: first to understand the tangled causal influences of environmental and social factors and second to demonstrate and refine the use of reproducible research pipelines to mitigate the reproducibility crisis. The papers included in the thesis demonstrate the successful application of the method of pipelines and each represent a new contribution to the state of our knowledge about the environmental exposures, in their social context. The results and conclusions of each paper will not be repeated here. Instead this section discusses the insights gained from the application of the pipelines method, noting the strengths and weaknesses of the approach. Potential future directions for applying the method of pipelines are then described. This discussion concludes by pointing toward the kind of eco-social epidemiological data analysis that will be more achievable (and credible) by using the methods of reproducible research pipelines.

6.1 Strengths

The work on the data analysis methods, and the data management approach has found that pipelines do indeed help create and analyse data and this helps to disentangle complex environmental and social causal pathways. The key attribute of working in a pipeline is to make available the analytic code and data (as appropriate)

along with journal papers. An important finding is that it is now easier than ever to write reports that incorporate the data and code as a compendium, using a computer scripting language, and to publish these as complete research products. The rationale for doing this is that it increases the credibility of the evidence.

6.2 Weaknesses

6.2.1 More time spent developing code instead of publishing results

One major weakness of the method is that it incurs an increased workload on the data analyst. In the first place an epidemiologist who seeks to engage in this practice of reproducible research must not only learn statistics and biology, but *also* undertake training in sometimes technical computer programming methods. Even after training, the practice of writing computer code takes more time than to conduct data analysis that is not reproducible. This is most obvious at the start of the project, or the first time an analysis method is attempted. In contrast it is also clear that the method saves time at later stages of the research. For example when iterations or revisions to the modelling and reports are needed, these can be made quickly and precisely by simply running the pipeline codes. Another example is that a new data analysis can be created using an old pipeline that can be re-designed for the new purpose.

6.2.2 Heterogeneous approaches to the organisation of material

There are many varied approaches that data analysts use for organisation of the material in a pipeline (including documentation, folders and files). This can make it hard to piece together and understand what has been done, even if the pipeline is able to recompute the results. This problem is mitigated to some extent because auditing the work is easier (Barnes 2010). Audits of data analysis are generally

only done by the analyst as part of their own quality assurance procedures. This is not necessarily something the readers will want to do, due to esoteric details and complicated methods, however it is increasingly asserted that the authors of code should be expected to make their work easier to review. For example in a recent editorial for the journal Nature Methods (Editorial 2015) it was pointed out:

> It could be argued that asking referees to review code and run software goes beyond the standards that have been applied in biology until now. After all, we do not ask referees to try to reproduce experimental results in their own labs. But the beauty of computational methods is precisely that this can be done. If we want robust analytical methods that generate believable and reproducible biological results, it should be done.

6.2.3 Sharing data is problematic

Weaknesses have been suggested by some scientists who state their concerns about sharing data in particular. Mills et al. (2015) outline a set of concerns from their work in ecology, arguing that data sharing will potentially have increased costs and risks including: increased misunderstanding of data, flawed science and decreased funding for data collection. They also proposed solutions to these concerns such as data copyright and access control, enhanced collaboration and communication, and improvements to data citation that would likely mitigate all these concerns. In the case of data used by epidemiologists any potentially re-identifiable health data may transgress privacy and confidentiality laws and so security of the computers the data are held on is of concern.

6.2.4 Being open risks cynical exploitation by those with malicious intentions

There is a salient risk that some scientists or lobby groups may cynically take advantage of minor errors that they find in shared code, to try to shame the scientist who made their work open.

In one example of this Leek (2015) reports how a reader:

> used our code (great), improved on it (great), identified a bug (great), and then did his level best to humiliate us both in front of the editor and the general public because of that bug (not so great).

Even in cases that are not so clearly vexatious, there is a tendency to demand excessively stringent standards for accepting claims. There is currently a debate in America about the 'Secret Science Reform Act of 2015', a bill that would 'prohibit the Environmental Protection Agency from proposing, finalizing, or disseminating regulations or assessments based upon science that is not transparent or reproducible' (Sarewitz 2015). Any delays in governments making decisive moves to implement effective actions to protect public health is clearly not a desirable outcome of the reproducible research methods, and is a violation of the precautionary principle (Oreskes & Conway 2013).

6.3 Future directions

6.3.1 Technological improvements that enhance the pipelines methods

There will continue to be developments of user friendly computer interfaces which will make it easier to write code, and easier to write prose that looks as good as it would had it been written in a traditional word processor program. In particular it has been found that collaborators who are not able to write code but want to contribute their input to the reporting (for example using 'track changes' in word documents) find this difficult. Additionally there will be a need for new tools to help analysts synthesise and integrate heterogenous data, while keeping their metadata in order, up to date, online and conforming to a standard schema. Software tools such as the Ecological Metadata Language (EML) interface for R (https://github.com/ropensci/EML) are maturing.

6.3.2 Improvements to understanding complex health problems

The future directions for studies of environmental changes in the context of strongly confounding social factors will be summed up next. One of the main areas of opportunity that have been identified is the use of a conceptual framework called the five capitals to guide the selection of explanatory variables for inclusion in the regression models.

The five capitals framework has been proposed as a method for disentangling such complex systems as these (Carney 1998; Scoones 1998; Ellis 2000; McMichael & Butler 2002; Nelson *et al.* 2005; Kokic *et al.* 2006). These five types of capital and their indicators are summarised as:

- ***Human capital***: indicated by the education and skill levels of the populace, as well as the historical background of settlement and growth in the locality; recent

population change and projections for the future; population characteristics such as labour force participation and occupation skill levels; and demographic composition.

- ***Financial capital***: indicators include the nature of core industries, population income and debt.
- ***Built capital***: includes public and private infrastructure (such as education facilities, banks, hospitals) and transport (such as airport access, rail access and density of the road network).
- ***Natural capital***: embodied in the geography, land, water and non-human eco-system of the regions.
- ***Social capital***: includes interaction between residents, organisations and services (including community groups, state and local government, business groups) as well as community values, traditions, cultural sites, religious involvement and the perceived desirability of living in the area.

For particular examples of how this approach might be implemented we turn to the problems of analysing the effect on suicide of drought. There are a large number of potential causal mechanisms that include drought, but also that would be considered extraneous to drought but important to include. For example, it has been found from the literature that there is evidence for an effect on suicide rates of the gun control intervention that was implemented in Australia in 1998 in NSW, around the same time that an extreme drought began (Leigh & Neill 2010). At that time it was also notable that the suicide rates during that drought were substantially reduced across the country (De Leo 2007), despite the postulated increased risk associated with drought.

An alternate example of potential variables that were not addressed was that of political regime which might be related to both drought and suicide. Conservative political regimes have been found to be statistically significantly related to increases in the population rate of suicide in NSW (Page *et al.* 2002) through mechanisms that are still poorly understood. It is postulated these may include a more selfish

and insular social milieu. It is possible that further analyses might reveal that the statistical association we found between the rainfall indicator and suicide is actually explained more completely by analysing causal intermediates, where suicide risk might increase alongside increased anomie during times of financial hardship, and when conservative political regimes are more likely to occur during droughts.

6.4 Conclusions

Environmental epidemiology faces fundamental questions of how to disentangle the effect of environmental change on human health from effects of social factors that might interact with these. These problems may lead to confounding of results of statistical data analysis. The studies reported in this thesis have advanced knowledge towards these goals. Yet questions remain. These are in part conceptual, in part definitional, in part statistical and in part due to insufficient quality of data. The problems of these four things impedes progress in this field; as for many other scientific endeavours. Advances in reproducibility address these four problems by enhancing transparency and allowing greater interactions between authors and readers. Publishing the data and code along with papers facilitates adequate communication of all assumptions, definitions, methods and data quality assurance checks.

The body of work presented in this thesis demonstrates the partial disentanglement of environmental and social variables, using sophisticated multi-variate statistical modelling. Such methods are valid and suitable to answer these questions, but results are hindered by the various difficulties in reproducibility which have been described. These obstacles can be reduced or even eliminated by adopting the strategies of reproducible research pipelines: an approach that produces publishable data and code along with papers by using computer scripting languages and tools. The findings of the published papers compiled in this thesis are thus open to close scrutiny by readers who may access the data and code. The data may not always be available for re-analysis (due to confidentiality or other reasons), however if an analyst has used the methods of pipelines, readers can have greater confidence that sensible and

rigorous work was done. This means that the data and methods steps were well organised and well documented. Beyond these issues are those also related to model selection, the suitability of such models and the assumptions used. Of note are those choices related to characterising exposure to potential causative agents, and the choice of health outcomes. These in particular must be documented, justified and have the theoretical underpinnings made explicit.

In conclusion, this thesis is not only about disentangling social and environmental factors. It is also about improving the ways that the details of how epidemiological studies were conducted can be communicated and interrogated. It demonstrates the use of the pipelines method to publish data and code along with papers. It is proposed that this method improves the rigour of the research and enables scientists to draw conclusions and interpretations of evidence that a raft of their peers would also make, if the methods are followed. For this all to occur, a much greater effort to record, archive and make data and code available is needed. The whole pipeline of data modelling and processing needs to be explicit, and then the conclusions drawn will be well founded.

Bibliography

Alberts, B., Cicerone, R.J., Fienberg, S.E., Kamb, A., McNutt, M., Nerem, R.M., Schekman, R., Shiffrin, R., Stodden, V., Suresh, S., Zuber, M.T., Pope, B.K. & Jamieson, K.H. (2015). Self-correction in science at work. *Science*, 348(6242), 1420–1422.

Albrecht, G., Sartore, G., Connor, L., Higginbotham, N., Freeman, S., Kelly, B., Stain, H., Tonna, A. & Pollard, G. (2007). Solastalgia: the distress caused by environmental change. *Australasian Psychiatry*, 15(S1), 95–98.

Australian Broadcasting Commission News. (2006). Drought lifts suicide rates: Kennett. http://www.abc.net.au/news/2006-10-13/drought-lifts-suicide-rates-kennett/1285734 [Accessed 7 Jun. 2012]

Australian National University. (2013). Guideline: Thesis by compilation. 1–3. https://policies.anu.edu.au/ppl/document/ANUP_003405 [Accessed 25 Nov. 2015]

Barnes, N. (2010). Publish your computer code: It is good enough. *Nature*, 467(7317), 753–753.

Bradford Hill, A. (1965). The environment and disease: Association or causation. *Proceedings of the Royal Society of Medicine*, 58: 295–300.

Brambor, T., Clark, W.R. & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14(1), 63–82.

Buck, S. (2015). Solving reproducibility. *Science*, 348(6242), 1403–1403.

Carney, D. (1998). *Implementing the Sustainable Rural Livelihoods Approach*. Department for International Development, London, UK.

Chamberlin, T. (1890). The method of multiple working hypotheses. *Science*, 39(2), 155–165.

Clancy, L., Goodman, P., Sinclair, H. & Dockery, D.W. (2002). Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study. *Lancet*, 360(9341), 1210–1214.

De Leo, D. (2007). Suicide mortality data need revision. *Medical Journal of Australia*, 186(3), 157–158.

Deisenhammer, E.A. (2003). Weather and suicide: The present state of knowledge on the association of meteorological factors with suicidal behaviour. *Acta Psychiatrica Scandinavica*, 108(6), 402–409.

Dixon, P. & Kalkstein, A. (2009). Climate-suicide relationships: A research problem in need of geographic methods and cross-disciplinary perspectives. *Geography Compass*, 3(6), 1–14.

Dockery, D.W., Rich, D.Q., Goodman, P.G., Clancy, L., Ohman-Strickland, P., George, P. & Kotlov, T. (2013). *Research Report on Mortality and Hospital Admissions in Ireland*. Health Effects Institute, Boston, USA.

Dominici, F., McDermott, A., Zeger, S.L. & Samet, J.M. (2002). On the use of Generalized Additive Models in time-series studies of air pollution and health. *American Journal of Epidemiology*, 156(3), 193–203.

Drinkel, E.E., Wu, L., Linden, A. & Dorta, R. (2014). Supporting information for: Synthesis, structure, and catalytic studies of palladium and platinum bis-sulfoxide complexes. *Organometallics*, 33

Editorial. (2015). Reviewing computational methods. *Nature Methods*, 12(12), 1099–1099.

El Kelish, A., Zhao, F., Heller, W., Durner, J., Winkler, J.B., Behrendt, H., Traidl-Hoffmann, C., Horres, R., Pfeifer, M., Frank, U. & Ernst, D. (2014). Ragweed (Ambrosia artemisiifolia) pollen allergenicity: SuperSAGE transcriptomic analysis upon elevated CO2 and drought stress. *BMC Plant Biology*, 14(1), 176.

Ellis, F. (2000). *Rural Livelihoods and Diversity in Developing Countries*. Oxford University Press, Oxford, UK.

Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge, United Kingdom.

Gelman, A. & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no 'fishing expedition' or 'p-hacking'. *Unpublished Draft*. http://www.stat.columbia.edu/ gelman/research/unpublished/ p_hacking.pdf [Accessed 10 Oct. 2015]

Gentleman, R. & Temple Lang, D. (2004). Statistical Analyses and Reproducible Research. *Journal of Computational and Graphical Statistics, Working Papers*, 16(1), 1–23.

Greenland, S., Pearl, J. & Robins, J.M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, 10(1), 37–48.

Guiney, R. (2012). Farming suicides during the Victorian drought: 2001-2007. *The Australian Journal of Rural Health*, 20(1), 11–15.

Hall, G.V., Hanigan, I.C., Dear, K.B.G. & Vally, H. (2011). The influence of weather on community gastroenteritis in Australia. *Epidemiology and Infection*, 139(6), 927–936.

Hanigan, I.C., Butler, C.D., Kokic, P.N. & Hutchinson, M.F. (2012). Suicide and drought in New South Wales, Australia, 1970-2007. *Proceedings of the National Academy of Sciences*, 109(35), 13950–13955.

Harley, D., Sleigh, A. & Ritchie, S. (2001). Ross River virus transmission, infection and disease: A cross-disciplinary review. *Clinical Microbiology Reviews*, 14(4), 909–932.

Health Effects Institute. (2003). *Special report: Revised analyses of time-series studies of air pollution and health*. Health Effects Institute, Boston, USA.

Holton, J., Pyle, J. & Curry, J. (2003). *Drought. Encyclopedia of Atmospheric Sciences*. Academic Press, New York, USA.

Irving, D. (2015). A minimum standard for publishing computational results in the weather and climate sciences. *Bulletin of the American Meteorological Society* (Epub ahead of print 2015). doi: 10.1175/BAMS-D-15-00010.1

Johnston, F.H., Hanigan, I.C., Henderson, S.B. & Morgan, G.G. (2013). Evaluation of interventions to reduce air pollution from biomass smoke on mortality in Launceston, Australia: retrospective analysis of daily mortality, 1994-2007. *BMJ*, 346(7890), e8446.

Johnston, F., Hanigan, I., Henderson, S., Morgan, G. & Bowman, D. (2011a). Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994-2007. *Environmental Research*, 111(6), 811–816.

Johnston, F.H., Hanigan, I.C., Henderson, S.B., Morgan, G.G., Portner, T., Williamson, G.J. & Bowman, D.M.J.S. (2011b). Creating an integrated historical record of extreme particulate air pollution events in Australian cities from 1994 to 2007. *Journal of the Air & Waste Management Association*, 61(4), 390–398.

Kim, Y., Kim, H. & Kim, D.-S. (2011). Association between daily environmental temperature and suicide mortality in Korea (2001-2005). *Psychiatry Research*, 186(2-3), 390–396.

King, G. (1995). Replication, replication. *PS: Political Science and Politics*, 28(3), 444–452.

Kokic, P., Davidson, A. & Boero Rodriguez, V. (2006). *Australia's Grains Industry: Factors Influencing Productivity Growth: ABARE Research Report*. ABARE

Research Report 06.22 prepared for the Grains Research; Development Corporation, Canberra, Australia.

Krieger, N. (1994). Epidemiology and the web of causation: Has anyone seen the spider? *Social Science & Medicine*, 39(7), 887–903.

Lambert, G., Reid, C., Kaye, D., Jennings, G. & Esler, M. (2003). Increased suicide rate in the middle-aged and its association with hours of sunlight. *American Journal of Psychiatry*, 160(4), 793–795.

Leek, J. (2015). De-weaponizing reproducibility. *Simply Statistics*. http://simplystatistics.org/2015/03/13/de-weaponizing-reproducibility/ [Accessed 7 Dec. 2015]

Leek, J.T. & Peng, R.D. (2015). Opinion: Reproducible research can still be wrong: Adopting a prevention approach. *Proceedings of the National Academy of Sciences of the United States of America*, 112(6), 1645–1646.

Leigh, A. & Neill, C. (2010). Do gun buybacks save lives? Evidence from panel data. *American Law and Economics Review*, 12(2), 462–508.

Loder, E. & Groves, T. (2015). The BMJ requires data sharing on request for all trials. *BMJ*, 350: h2373.

Lucas, R.M. & McMichael, A.J. (2005). Association or causation: Evaluating links between 'environment and disease'. *Bulletin of the World Health Organization*, 83(10), 792–795.

Lumley, T. & Sheppard, L. (2003). Time series analysis of air pollution and health: Straining at gnatts and swallowing camels? *Epidemiology*, 14(1), 13–14.

Martin, K.L., Hanigan, I.C., Morgan, G.G., Henderson, S.B. & Johnston, F.H. (2013). Air pollution from bushfires and their association with hospital admissions in Sydney, Newcastle and Wollongong, Australia 1994-2007. *Australian and New Zealand Journal of Public Health*, 37(3), 238–243.

Mazdiyasni, O. & AghaKouchak, A. (2015). Substantial increase in concurrent droughts and heatwaves in the United States. *Proceedings of the National Academy of Sciences*, 2015(24), 201422945.

McMichael, A.J. (2013). Impediments to comprehensive research on climate change and health. *International Journal of Environmental Research and Public Health*, 10(11), 6096–6105.

McMichael, A.J. (1999). Prisoners of the proximate. *American Journal of Epidemiology*, 149(10), 887–897.

McMichael, A.J. & Butler, C.D. (2002). Global health trends: Evidence for and against sustainable progress. *International Union for the Scientific Study of Popula-*

tion Committee on Emerging Health Threats. http://www.demogr.mpg.de/papers/workshops/020619_paper25.pdf [Accessed 21 Sep. 2003]

McMichael, A.J., Butler, C.D. & Dixon, J. (2015). Climate change, food systems and population health risks in their eco-social context. *Public Health* (Epub ahead of print April 2015). doi: 10.1016/j.puhe.2014.11.013

Meade, M. & Earickson, R. (2000). *Medical Geography.* Guilford Press, New York, USA.

Millennium Assessment. (2003). *Ecosystems and Human Well-Being: A Framework for Assessment. Chapter 2: Ecosystems and Their Services.* Island Press, Washington DC, USA.

Mills, J.A., Teplitsky, C., Arroyo, B., Charmantier, A., Becker, P.H., Birkhead, T.R., Bize, P., Blumstein, D.T., Bonenfant, C., Boutin, S., Bushuev, A., Cam, E., Cockburn, A., Cté, S.D., Coulson, J.C., Daunt, F., Dingemanse, N.J., Doligez, B., Drummond, H., Espie, R.H.M., Festa-bianchet, M., Frentiu, F., Fitzpatrick, J.W., Furness, R.W., Garant, D., Gauthier, G., Grant, P.R., Griesser, M., Gustafsson, L., Hansson, B., Harris, M.P., Jiguet, F., Kjellander, P., Korpimki, E., Low, M., Mcadam, A., Margalida, A., Meril, J., Mller, A.P., Nakagawa, S., Oro, D., Prt, T., Pelletier, F., Potti, J., Ropert-coudert, Y., Roulin, A., Sedinger, J.S., Swenson, J.E., Thébaud, C., Visser, M.E., Wanless, S., Westneat, D.F. & Wilson, A.J. (2015). Archiving primary data : Solutions for long-term studies. *Trends in Ecology & Evolution*, 30(10), 581–589.

Naeher, L.P., Brauer, M., Lipsett, M., Zelikoff, J.T., Simpson, C.D., Koenig, J.Q. & Smith, K.R. (2007). Woodsmoke health effects: A review. *Inhalation Toxicology*, 19(1), 67–106.

Nelson, R., Kokic, P., Elliston, L. & King, J. (2005). Structural adjustment. *Australian Commodities*, 12(1), 171–179.

Nicholls, N., Butler, C.D. & Hanigan, I. (2006). Inter-annual rainfall variations and suicide in New South Wales, Australia, 1964-2001. *International Journal of Biometeorology*, 50(3), 139–143.

Nosek, B.A. & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, 45(3), 137–141.

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E.L., Simonsohn, U., Soderberg, C., Spellman, B.A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E.J., Wilson, R. & Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425.

Nuzzo, R. (2015). How scientists fool themselves – and how they can stop. *Nature*, 526(7572), 182–185.

O'Brien, L.V., Berry, H.L., Coleman, C. & Hanigan, I.C. (2014). Drought as a mental health exposure. *Environmental Research*, 131: 181–187.

Oliver, R. & Hetzel, B. (1973). An analysis of recent trends in suicide rates in Australia. *International Journal of Epidemiology*, 2(1), 91.

Oreskes, N. & Conway, E.E.M. (2013). The Collapse of Western Civilization: A View from the Future. *Daedalus*, 142(1), 40–58.

Page, A., Morrell, S. & Taylor, R. (2002). Suicide and political regime in New South Wales and Australia during the 20th century. *Journal of Epidemiology and Community Health*, 56(10), 766–772.

Pavlovskii, E. (1966). *Natural Nidality of Transmissible Diseases*. University of Illinois Press, Champaign, USA.

Pearl, J., Glymour, M. & Jewell, N.P. (2016). Chapter 1: Preliminaries: Statistical and Causal Models. In *Causal Inference in Statistics: A Primer.*, pp. 1–34. John Wiley & Sons, West Sussex, United Kingdom.

Peng, R.D. (2009). Reproducible research and Biostatistics. *Biostatistics*, 10(3), 405–408.

Peng, R.D. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226–1227.

Peng, R.D. (2015). The reproducibility crisis in science: A statistical counterattack. *Significance*, 12(3), 30–32.

Peng, R.D. & Dominici, F. (2008). *Statistical Methods for Environmental Epidemiology with R. A Case Study in Air Pollution and Health*. Springer Science & Business Media, New York, USA.

Peng, R.D. & Eckel, S.P. (2009). Distributed Reproducible Research Using Cached Computations. *Computing in Science & Engineering*, 11(1), 28–34.

Peng, R.D., Dominici, F. & Zeger, S.L. (2006). Reproducible epidemiologic research. *American Journal of Epidemiology*, 163(9), 783–789.

Platt, J.R. (1964). Strong inference. *Science*, 146(3642), 347–353.

Pope, C.A. & Dockery, D.W. (2006). Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air & Waste Management Association*, 56(6), 709–742.

Postolache, T.T., Stiller, J.W., Herrell, R., Goldstein, M.A., Shreeram, S.S., Zebrak, R., Thrower, C.M., Volkov, J., No, M.J., Volkov, I., Rohan, K.J., Redditt, J., Parmar,

M., Mohyuddin, F., Olsen, C., Moca, M., Tonelli, L.H., Merikangas, K. & Komarow, H.D. (2005). Tree pollen peaks are associated with increased nonviolent suicide in women. *Molecular Psychiatry*, 10(3), 232–235.

Powers, J.R., Dobson, A.J., Berry, H.L., Graves, A.M., Hanigan, I.C. & Loxton, D. (2015). Lack of association between drought and mental health in a cohort of 45-61 year old rural Australian women. *Australian and New Zealand Journal of Public Health*, 39(6), 518–23.

Powers, J.R., Loxton, D., Baker, J., Rich, J.L. & Dobson, A.J. (2012). Empirical evidence suggests adverse climate events have not affected Australian women's health and well-being. *Australian and New Zealand Journal of Public Health*, 36(5), 452–457.

Preti, A. (1998). The influence of climate on suicidal behaviour in Italy. *Psychiatry Research*, 78(1-2), 9–19.

Reid, C.E., Snowden, J.M., Kontgis, C. & Tager, I.B. (2012). The role of ambient ozone in epidemiologic studies of heat-related mortality. *Environmental Health Perspectives*, 120(12), 1627–1630.

Rothman, K.J. (1976). Causes. *Journal of Epidemiology*, 104(6), 587–592.

Sólymos, P. & Fehér, Z. (2008). The mefa package: a tool for reproducible data processing in biogeography. *International Biogeography Society Newsletter*. http://biogeography.blogspot.com.au/2008/04/mefa-package-tool-for-reproducible-data.html [Accessed 10 Aug. 2015]

Sarewitz, D. (2015). Reproducibility will not cure what ails science. *Nature*, 525(7568), 159–159.

Schulte, E., Davison, D., Dye, T. & Dominik, C. (2012). A multi-language computing environment for literate programming and reproducible research. *Journal of Statistical Software*, 46(3).

Scoones, I. (1998). *Sustainable Rural Livelihoods: A Framework for Analysis. Working Paper 72*. Institute of Development Studies, Brighton, UK.

Smith, K. (1993). Fuel combustion, air pollution exposure, and health: The situation in developing countries. *Annual Review of Energy and the Environment*, 18: 529–66.

Smith, D.I., Hutchinson, M.F. & McArthur, R.J. (1992). *Climatic and Agricultural Drought: Payments and Policy*. Australian National University, Canberra, Australia.

Speldewinde, P., Cook, A., Davies, P. & Weinstein, P. (2009). A relationship between environmental degradation and mental health in rural Western Australia. *Health and Place*, 15: 880–887.

Stanke, C., Kerac, M., Prudhomme, C., Medlock, J. & Murray, V. (2013). Health effects of drought: A systematic review of the evidence. *PLoS Currents Disasters*,

1(1), 1–38.

Vally, H., Peel, M., Dowse, G.K., Cameron, S., Codde, J.P., Hanigan, I. & Lindsay, M.D. a. (2012). Geographic information systems used to describe the link between the risk of Ross River virus infection and proximity to the Leschenault Estuary, WA. *Australian and New Zealand Journal of Public Health*, 36(3), 229–235.

Vins, H., Bell, J., Saha, S. & Hess, J. (2015). The mental health outcomes of drought: A systematic review and causal process diagram. *International Journal of Environmental Research and Public Health*, 12(10), 13251–13275.

Wilson, E. (1984). *Biophilia*. Harvard University Press, Cambridge, USA.

Wilson, L.A., Morgan, G., Hanigan, I.C., Johnston, F., Abu-Rayya, H., Broome, R., Gaskin, C. & Jalaludin, B. (2013). The impact of heat on mortality and morbidity in the Greater Metropolitan Sydney Region: a case crossover analysis. *Environmental Health : a Global Access Science Source*, 12(1), 98.

Xie, L., Weichel, B., Ohm, J. & Zhang, K. (2011). An integrative analysis of DNA methylation and RNA-Seq data for human heart, kidney and liver. *BMC Systems Biology*, 5(Suppl 3), S4.

Appendix 1: Details of research outputs

Table A1.1: Details of research outputs

Publication type	Details
Paper 1	Hanigan, I.C., Butler, C.D., Kokic, P.N. and Hutchinson, M.F. (2012). Suicide and drought in New South Wales, Australia, 1970-2007. Proceedings of the National Academy of Sciences of the United States of America, 109(35), 13950-13955. doi:10.1073/pnas.1112965109
Conference presentation and working paper	Hanigan, I.C., Fisher, D., and McEachern, S. (2013). Open software - restricted data: a suicide / climate case study. In National Climate Change Adaptation Research Facility Conference Climate Adaptation knowledge and partnership. June 2013, Sydney.
Dataset 1	Hanigan, I. (2012). Monthly drought data for Australia 1890-2008 using the Hutchinson Drought Index. The Australian National University Australian Data Archive. doi:10.4225/13/50BBFD7E6727A
Software package 1	Hanigan, I., Porfirio, L. and Hutchinson, M. (2012). The Hutchinson Drought Index Algorithm. https://github.com/ivanhanigan/HutchinsonDroughtIndex.
Dataset 2	Hanigan, I. C. (2015). Woodland Restoration Plot Network: Climatic Drought Index Data, Western Sydney Parklands (Western Sydney Regional Park), Australia, 1887 - 2014. Long Term Ecological Research Network (LTERN).
Paper 2	O'Brien, L. V., Berry, H. L., Coleman, C., and Hanigan, I. C. (2014). Drought as a mental health exposure. Environmental Research, 131: 181-187. doi:10.1016/j.envres.2014.03.014
Paper 3	Powers, J. R., Dobson, A. J., Berry, H. L., Graves, A. M., Hanigan, I. C., and Loxton, D. (2015). Lack of association between drought and mental health in a cohort of 45-61 year old rural Australian women. Australian and New Zealand Journal of Public Health. 39(6):518-23.
Paper 4	Johnston, F., Hanigan, I., Henderson, S., Morgan, G., Portner, T., Williamson, G., and Bowman, D. (2011). Creating an integrated historical record of extreme particulate air pollution events in Australian cities from 1994 to 2007. Journal of the Air Waste Management Association, 61(4), 390-398. doi:10.3155/1047-3289.61.4.390

Table A1.1: Details of research outputs

Publication type	Details
Dataset 3	Hanigan, I. C. (2015). The Validated Bushfire Smoke Events Database. https://gislibrary-extreme-weather.anu.edu.au/biomass_smoke_events
Software package 2	Hanigan, I. C. (2015). Biosmoke Validated Events. https://github.com/swish-climate-impact-assessment/BiosmokeValidatedEvents.
Paper 5	Johnston, F.J., Hanigan, I.C., Henderson, S.B., Morgan, G.G, and Bowman, D.M.J.S. (2011). Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994-2007. Environmental Research 111 (6): 811-816. doi:10.1016/j.envres.2011.05.007
Paper 6	Martin, K. L., Hanigan, I. C., Morgan, G. G., Henderson, S. B., and Johnston, F. H. (2013). Air pollution from bushfires and their association with hospital admissions in Sydney, Newcastle and Wollongong, Australia 1994-2007. Australian and New Zealand Journal of Public Health, 37(3), 238-243. doi:10.1111/1753-6405.12065
Paper 7	Wilson, L. A., Morgan, G., Hanigan, I. C., Johnston, F., Abu-Rayya, H., Broome, R., Gaskin, C., and Jalaludin, B. (2013). The impact of heat on mortality and morbidity in the Greater Metropolitan Sydney Region: a case crossover analysis. Environmental Health : A Global Access Science Source, 12(1), 98. doi:10.1186/1476-069X-12-98
Paper 8	Johnston, F.H., Hanigan, I.C., Henderson, S.B., and Morgan, G.G. (2013). Evaluation of interventions to reduce air pollution from biomass smoke on mortality in Launceston, Australia: retrospective analysis of daily mortality, 1994-2007. BMJ, 346, e8446. doi:10.1136/bmj.e8446
Paper 9	Hall, G. V, Hanigan, I. C., Dear, K. B. G., and Vally, H. (2011). The influence of weather on community gastroenteritis in Australia. Epidemiology and Infection, 139(6), 927936. doi:10.1017/S0950268810001901
Dataset 4	Hanigan, I. (2010). Meteorological Data for Australian Postal Areas. Australian Data Archive. doi:10.4225/13/50BBFCFE08A12

Table A1.1: Details of research outputs

Publication type	Details
Paper 10	Vally, H., Peel, M., Dowse, G. K., Cameron, S., Codde, J. P., Hanigan, I., and Lindsay, M. D. A. (2012). Geographic Information Systems used to describe the link between the risk of Ross River virus infection and proximity to the Leschenault Estuary, WA. Australian and New Zealand Journal of Public Health, 36(3), 229235. doi:10.1111/j.1753-6405.2012.00869.x
Software package 3	Hanigan, I.C. (2015). Reproducible code for Vally et al. 2012 Ross River Rates. https://github.com/ivanhanigan/Vally2012-RossRiverRates

Appendix 2:

Data management procedures for reproducible research pipelines

Ivan C. Hanigan

Abstract

This unpublished working paper was written to accompany the material included in the PhD thesis 'Using Reproducible Research Pipelines to Help Disentangle Health Effects of Environmental Changes from Social Factors'. It sets out the key data management and analysis principles that were found to be most effective for the reproducible synthesis and integration of heterogeneous datasets for analysis and reporting. The draft was last updated August 21, 2016. The version submitted with the thesis is available on the phd_appendix branch of the Github repository: https://github.com/swish-climate-impact-assessment/swish_data_management_procedures.

Contents

1 Introduction

There is a need for developing an evidence-based set of best practice guidelines for data management procedures that support reproducibility in all fields of computational data analysis (Long, 2008; Noble, 2009; Peng, 2015). Reproducibility is the ability to recompute the results of a data analysis with the original data (as distinct from replication which involves analysing independently collected data (Peng, 2011)). The examples drawn together in this report come from experiences and use-cases found from implementing reproducible research pipelines in an eco-social epidemiological research context. This emerging paradigm mixes environmental and social epidemiology and is inherently concerned with complex systems. To do this work integration of heterogeneous data sources, and synthesising new datasets, is required. Then analyses that aim to recognise subtle and complicated patterns in the environmental and social determinants of health must be rigorously and transparently conducted (McMichael, 2013). This document outlines a suite of data management procedures that have been found to effectively assist the development of reproducible research pipelines in this context.

1.1 The 'reproducibility crisis'

All data analyses are reproducible to a varying degree of difficulty. A data analysis might be reproducible but require thousands of hours of work. A primary challenge for reproducible data analysis is to make analyses that are *easy* to reproduce.

In essence this requires attention to be turned to the issue of how the data and analytical steps amassed – toward a reality where this is archived and there is a good understanding all round as to how the study were set up and conducted. Different assumptions or different treatment of the data could conceivably lead to different inferences and conclusions being drawn, such as in the example shown by Silberzahn and Uhlmann (2015) in which 29 research teams were given the same dataset but reached a wide variety of conclusions using different methods on the same dataset to answer the same question.

This is partly because of an underlying complexity in the information drawn from complex systems involving multi-causality, and partly because of different assumptions and different backgrounds and viewpoints. A finding that a variable does or does not cause a disease, might be drawn honestly from the same set of data.

1.2 A common (flawed) approach for generating statistical reports

A common approach (with inherent flaws that make it error-prone) was identified by Scott (2010), and the examples are paraphrased here. First: the data entry, cleaning, preparation and possibly statistical analyses are conducted by 'point-and-click' procedures using spreadsheet software such as Microsoft (MS) Excel. This introduces well-known issues with handling of dates, coloured cells, hidden columns, missing data, poor algorithms and unreliable results (McCullough and Heiser, 2008). In cases where data are imported to a program such as STATA or SPSS, further point-and-click data preparation and statistical analyses often occur. Post-analysis, spreadsheets such as MS Excel are regularly used to record or format the desired results, and generate figures. Finally, the results (text, tables and figures) from the data analysis system are inserted into a word processor (e.g. MS Word) using 'copy-and-paste' procedures (or typed by hand).

Problems with this common, flawed approach according to Scott (2010) are:

- You sit down to finish writing your manuscript. You realize that you need to clarify one result by running an additional analysis. You first re-run the primary analysis. Major problem: the primary results don't match what you have in your paper.
- When you go to your project folder to run the additional analysis, you find multiple data files, multiple analysis files, and multiple results files. You can't remember which ones are pertinent.
- You have just spent the week running your analysis and creating a results report (including tables and figures) to present to your collaborators. You then receive an email from your boss asking you to regenerate the report based on a subset of the original data set and including an additional set of analyses – she would like it by tomorrow's meeting.
- With point and click programs there is no way to record the steps performed that generated the documented results.
- Common to keep analysis code, results, and reports as separate files and to save various versions of each of these as separate files. After several modifications of one or more of the files involved, becomes unclear which version of the files exactly correspond to the desired analysis and results.
- Every time analyses and/or results change, have to regenerate the results report by hand – very time consuming.
- Very easy for human error to creep into results report (e.g., typing in results by hand, copying/pasting the wrong tables/figures).

1.3 Reproducible research reports: A better alternative

It is widely recommended that a better approach is to create Reproducible Research Reports (RRR) (Healy, 2013). This embeds the analysis into the report so that the code to clean and prepare the data or to perform the desired statistical analysis is included in the document that contains the documentation and text of the report. Solutions have been developed that combine both the data analysis code and the descriptive prose that constitutes the publishable report into a compendium (Gentleman and Temple Lang, 2004; Schulte et al., 2012).

RRR are written using a scripting language for statistical computing and graphics. An example using the R language is presented in box 1 below. The report is made up of ordinary text written in a suitable format that enables the computational process to recognise it as text. An example is the Markdown format which is very similar to text used when authoring word processor documents (http://rmarkdown.rstudio.com). There are also chunks of pure statistical programming code (such as R codes) that perform data manipulations and analyses when the document is 'evaluated'. When the processing stage is run a report document is generated that includes both content as well as the output of any embedded computer code 'chunks' within the document. These are distinguished from the regular text by a special delimiter at their beginning and end.

Box 1: An example Reproducible Research Report using R and markdown

````
---
title: "Reproducible report example"
author: "Ivan C. Hanigan"
output:  pdf_document
---
# Some exploratory analysis
In this section we do some exploratory analysis of the NMMAPS data for
deaths in Chicago 1987-2000. The code, messages and intermediary
results are hidden in the resulting report document.
```{r, echo = FALSE, message = FALSE}
## We begin  by reading in the data file:
## If using our own data we would use 'read.csv' or a similar tool to import data to R
# my.data <- read.csv('data/sampledata.csv',header=TRUE)
## for this example use data that are included in the dlnm package
library(dlnm)
# look at the structure of the data
# str(chicagoNMMAPS)
# summary(chicagoNMMAPS)
```
We made a simple scatter plot shown below
```{r, echo = FALSE, message = FALSE}
## make some plots. first by day
# with(chicagoNMMAPS, plot(date, cvd, type = "l"))
# we suspect a relationship between temperature and deaths
with(chicagoNMMAPS, plot(temp, cvd, pch = 16, cex = .6))
title(main = "A scatter plot of daily temperatures against deaths")
```
We ran some exploratory models. A Poisson GAM with smooth functions
on temperature and time was compared to a linear fit on temperature.
```{r, echo = FALSE, message = FALSE}
## GAM model with a smooth function to estimate the temperature effect
# and another to control for time trends
library(mgcv)
fit1 <- gam(cvd ~ s(temp) + s(time), data=chicagoNMMAPS, family = "poisson")
# we can access post-estimation summary statistics
# summary(fit1)
# do model testing to confirm that the error terms are distributed as assumed
# gam.check(fit1)
# or just plot the exposure-response function
plot(fit1, select = 1)
title(main = "The exposure-response function estimated using MGCV")
aic1 <- AIC(fit1)
# compare this to a linear term for temperature
aic0 <- AIC(gam(cvd ~ temp + s(time), data=chicagoNMMAPS, family = "poisson"))
# calculate the delta aic
aici <- aic1 - aic0
```
The result can be automatically inserted to the text. This model has
a delta AIC of `r round(aici,1)` (smoothed minus linear term).
````

The result can be seen at this link (https://github.com/swish-climate-impact-assessment/swish_data_management_procedures/blob/phd_appendix/Reproducible_report_example.pdf). In this example the Markdown engine is used to construct the final report by 'weaving' together the prose and the code. The prose is written in 'markdown' which is a simple way to use 'markup' commands to tell the program to do formatting on the inputs. For example the first 'hash' (#) symbol tells the program that this line should be written in the style of heading-1. The three 'back tick' marks (') tell the program that the following text should be interpreted as R code. Inside these 'chunks' the 'hash' symbol is interpreted as a comment and the line is not executed.

If an analysis is published as a reproducible report, readers can have greater confidence in the work that was done, and verify this for themselves if questions remain. However, it has also been recognised that reproducible research can still be wrong and a prevention approach has been recommended that incorporates evidence-based data analysis tools and techniques into such a pipeline (Leek and Peng, 2015a).

1.4 Reproducible Research Pipelines (RRP) defined

The techniques of pipelines described here are targeting the integrity of the process of data selection, the robustness and suitableness of the methods used, a commonsense and well-argued selection of health outcomes and environmental or social exposures, and the clarity and transparency of the methods used.

To achieve this, a guiding principle is that analysts should effectively implement 'pipelines' of method steps and tools. Standardised and evidence-based methods based on conventions developed from many data analysts approaching the problems in a similar way should be used, rather than each analyst configuring a pipeline to suit particular individual or domain-specific preferences (Borer et al., 2009; White et al., 2013).

Noble (2009) points out that 'the principles behind organizing and documenting computational experiments are often learned on the fly, and this learning is strongly influenced by 'personal predilections'. Leek and Peng (2015b) describe this as data analysis being 'taught through an apprenticeship model, and different disciplines develop their own analysis subcultures'. By codifying what an appropriate pipeline would contain, data analysis will be more robust. According to Peng (2015), there should not be a 'lonely data analyst' coming up with their own method. If a researcher conducted an analysis using an evidence-based reproducible research pipeline 'you could at least have a sense that something reasonable was done' (Peng, 2015).

1.5 The core components of a pipeline

The core concepts and flow of steps in the method are shown in Figure 1 (after Peng et al. (2006) and Sólymos and Fehér (2008)). In this model there are two main actors: the author and the reader. The author moves from left to right, from initial hypothesis and study design, through data collection and pre-processing, to analysis and reporting. The aim is to conduct all steps of the analysis work in such a way that the key dataset and code script can be sent through the distribution mechanism and the reader can easily move from right to left. Thus the reader can start with the published results and then dig deeper by assessing the analysis code and analytic data to gain full understanding of the methods steps.

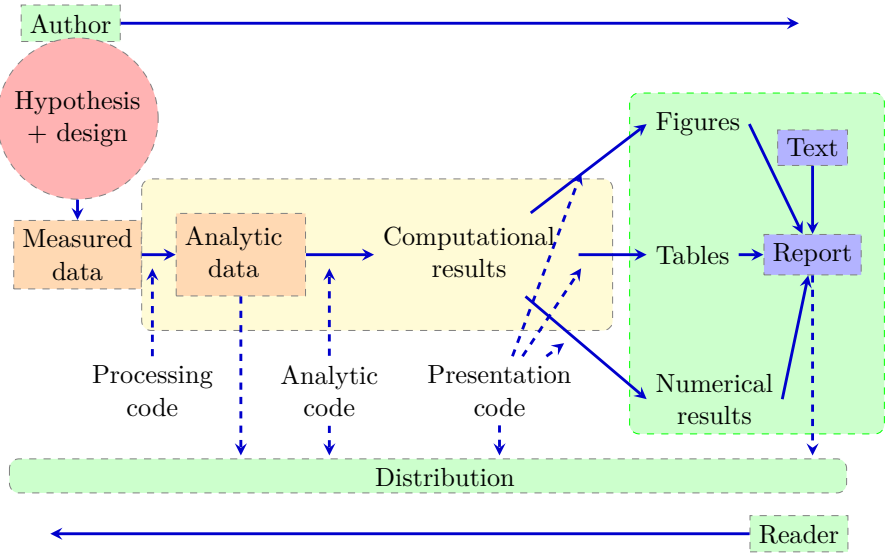


Figure 1: A schematic diagram representing the reproducible research pipeline

Peng et al. (2006) distilled a core set of components for reproducibility from earlier work including that of Schwab et al. (2000). These are:

- Hypothesis and design
- Data (measurement, pre-processing, analytic)
- Analysis Methods
- Documentation (of all steps)
- Distribution (of the paper, data and code).

1.5.1 Hypothesis and design

The first stage of the pipeline is hypothesis generation and study design. In this stage documentation should explain the literature base supporting the study, the decisions made in selection of explanatory factors for inclusion, decisions made such as the experimental unit, observational unit, measurement method, as well as spatial or temporal extent. This information will also be needed for ethical review and approval.

1.5.2 Data

The data that were measured should be well managed, however the requirements for accessing the original raw data are less important than for the analytical dataset. Descriptions of how the measured data were transformed into the analytic data should be available. Public data repositories or institutional services such as university libraries should be used to ensure longevity of the data storage.

1.5.3 Methods

The software code underlying the principal results needs to be made available. In addition, the computer environment necessary to execute that code should be described adequately to 'deploy' a new computer set-up that can reproduce the computations needed.

1.5.4 Documentation

Adequate documentation of the code and data should be available to enable others to repeat the analyses and to conduct other similar ones. This can take the form of metadata, reports, journal papers or even books (Peng and Dominici, 2008). Indeed textbooks on statistical methods can benefit greatly from being accompanied by data and analytical code to enhance their pedagogic functions (Barnett et al., 2014; Barnett and Dobson, 2010).

Inadequate documentation can lead to data misuse. This may be due either to honest misunderstandings about data attributes (no dataset is perfect and self-explanatory, see Michener et al. (1997)) or intentional misuse for malicious or selfish reasons (for example the misuse of data by Bjorn Lomborg to support the argument that environmental health conditions are actually improving. See Bodnar et al. (2004) for a discussion on Lomborgś misuse of data). There have also been notable examples of mistakes in data analyses used for climate change science. See Cai et al. (2010) for a discussion of one such case. The careful storage and curation of datasets is also critical because data from many studies are lost (Pullin and Salafsky, 2010; Vines et al., 2014).

An important underpinning to reproducible research is the reproducible report. This is the ultimate form of documentation because the information that represents the outputs of the research is written alongside the code that performs the computations that are being described. There has been many recent advances made in terms of tools for reproducible reports such as R markdown and knitr (Xie, 2014).

Metadata should be created and maintained as a priority task at all stages of the data analysis process. An international standard should be preferred over selectively choosing what information one collects and what fieldnames one uses to describe each item of documentation. Ecological Metadata Language (EML) and the Data Documentation Inititative (DDI) are two such standards that offer useful semantic constructs for describing epidemiological data.

Time and effort may be saved by considering metadata requirements at the commencement of a study, rather than trying to recall all the details later. If metadata adheres to a standard schema, it can be used in catalogues to enable fast searching and retrieval, or in machine-to-machine data queries that assist data access and use.

1.5.5 Distribution

Distribution or dissemination of the material needs to use a standard method if they are to be used by others. It is not enough just to provide access to the software and data, but also adequate documentation is required to explain and potentially assist downstream users to piece these together.

2 Principles for organising projects, datasets and files

For data to be reused in the future, files and related documents need to be carefully managed to allow future users (including the original collector) to find and understand them. A formalised approach to data management should be developed, and following a widely agreed 'standard' if possible. This example shows the use of the Ecological Metadata Language (EML) concepts of Projects, Datasets and Entities. EML was developed primarily for creating metadata and allows sufficient detail to describe the collection process and record decisions that were made during the creation of the data. In EML the elements of any dataset can be seen as a nested hierarchy at three levels shown in figure 2.

1. The Project level: this is an overarching grouping of data. It might be indicative of the principal investigator or organisation who provided the data, or a programme of research studies (sub-projects).
2. The Dataset level: this is a distinct grouping of data that might be organised around a particular time period or geographical region.
3. The Entity level: This grouping of data includes data files (such as tables in CSV or Excel, shape-files and raster images) or documents (such as metadata descriptions or related publications).

This conceptual framework can be very useful for the organisation of the work constituting a single pipeline, as well as when working with multiple pipelines within several projects.

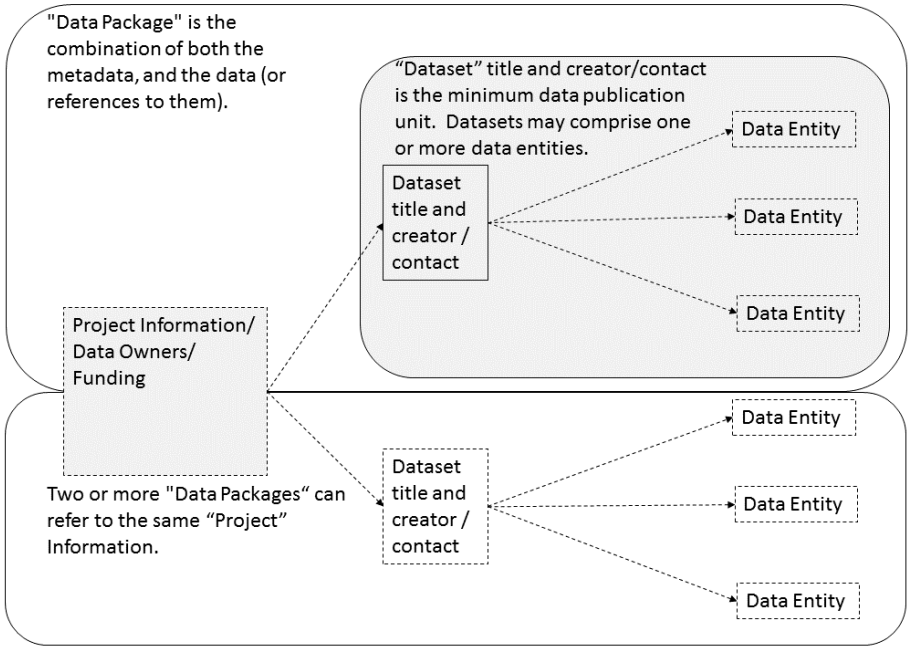


Figure 2: The EML approach to managing Projects, Datasets and Entities

2.1 Procedures when conducting a reproducible research analysis

Having defined above the principle components for a pipeline there are procedural questions about how to go about compiling those. The key steps include:

- data management plans and data inventories
- planning and implementing a pipeline
- tracking method steps.

These three topics will be explored in each of the next three sections of this report.

3 Data management plan and data inventory

In eco-social epidemiology there is a need for a data management plan and a data inventory that enables individual scientists, or multidisciplinary teams of scientists, to manage large and heterogeneous collections of disparate data sources efficiently. Keeping track of all the elements of a linked health, social and environmental database is very challenging, despite major improvements in data management software, web-portals and virtual laboratories (Fleming et al., 2014).

Effective data management policies and procedures are essential in managing data-related risk. Such risks include data loss or corruption, technological obsolescence, breaches of privacy or copyright, and errors or misuse.

Data management plans are needed for developing procedures and processes to keep data safe. There is an issue when ensuring that all relevant data are collected in deciding what is relevant. Keeping an up-to-date data inventory and careful organisation of all folders and files helps mitigate these problems.

Whether data management is the responsibility of the individuals collecting or collating it, or of the lead scientist, clarity on how and where data are stored and who manages it is vital, as is a 'succession plan' that sets out the vision of the data collections preservation and re-use into the future.

3.1 Data storage and access

Some datasets such as sensitive personal information about suicide or climate change scenarios with restrictions due to privacy and confidentiality rules, or because of protected intellectual property, need to be accessed in a restricted way. This complicates the implementation of the method of pipelines which dictates that all the steps, models and assumptions need to be made transparent and available for scientific debate even though the datasets may require authorisation to access. Restrictions around access to data have increased recently in Australia. As an example the custodians of the national mortality database made it virtually impossible to access these data for several years after the discovery of an incident in which Australian population health researcher Dr Stephen Begg was reported to have hacked into the database in an illegal act (O'Keefe, 2007). The subsequent investigation by the data custodians led to a wide ranging modification to the procedures for approval and provision of these data that make the access much more restricted. Appropriate access to data is therefore required to address this issue. In the work reported in the conference presentation in this thesis, a range of available workflow tools for data management and analysis were investigated and developed.

The key components of the data management system as identified above are shown graphically in figure 3. A project can contain many datasets (folders), which in turn may hold many entities (files). This organisation model for a project can be described in terms of 'general' projects (about general contextual data, accessible to the entire group) or 'specific' projects (about a specific Health/Exposure study and will have a specific subset of authorised people who can access it). General data can go into the main storage folder however specific data needs to have more secure storage.

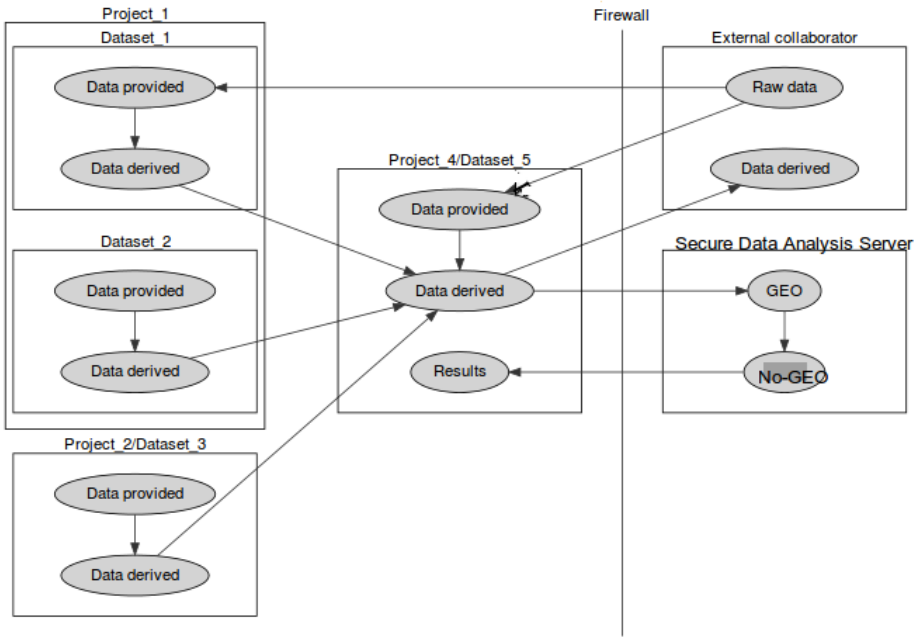


Figure 3: A schematic diagram of the management of large multi-institute collaborative project

3.2 Case study 1: EML and folder structure

In figure 4 the conceptual framework described above is implemented in a standardised folder structure. The main storage location for this data collection is called 'Research data'. This computer drive is then structured in a simple hierarchy of projects (folders), datasets (sub-folders) and entities (sub-sub-folders or individual files). Entities may be individual files, or groups of files in a sub-sub-folder to cater for data structures such as those where there are a collections of files that make up a single entity such as the Shapefile or Raster Image dataset as used in Geographical Information Systems (GIS) software. As seen in figure 3 it might make sense to group all entities into folders that delineate those files provided as raw data and those that were derived by some process within the project.

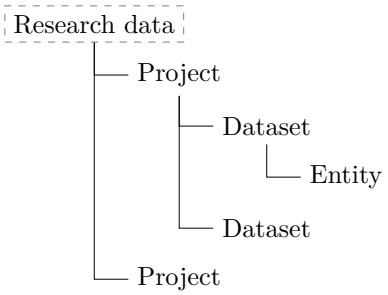


Figure 4: Conceptual framework for grouping data files (entities) within datasets and projects

4 Planning and implementing a pipeline

It can be much easier to conceptualise a complicated data analysis method than to implement this as a reproducible research pipeline. The most effective way to implement a pipeline is by methodically tracking each of the steps taken, the data inputs needed and all the outputs of the step. If done in a disciplined way then the analyst or some other person could 'audit' the procedure easily and access the details of the pipeline they need to scrutinise.

4.1 A standardised data analysis pipeline framework

One method that was selected for use in the papers of this thesis was the concept of the Load-Clean-Functions-Do (LCFD) framework. This was first proposed by Josh Reich on the open-source software discussion forum called 'stack overflow' (http://stackoverflow.com/a/1434424), and then encoded into the 'makeProject' R package (http://cran.r-project.org/web/packages/makeProject/makeProject.pdf). The approach is demonstrated in case study 2 below.

4.2 Case study 2: Simple pipeline using the makeProject package

```
# in an interactive R session at the command line choose your project directory
setwd("~/projects")
# load the required functions from the makeProject package
library(makeProject)
# use the makeProject function to
makeProject("my_first_pipelines_project")

### gives
/my_first_pipelines_project/
    /code/**.R
    /data/
    /DESCRIPTION
    /main.R

# in main.R you put these lines into the script
# and run them as the steps of the pipeline evolve
source("code/load.R")
source("code/clean.R")
source("code/func.R")
source("code/do.R")

# Reporting is then a matter of choice
## If using the rmarkdown approach there would be an Rmd file that contained the prose
## and turned into a PDF, HTML or Word document with a line such as
rmarkdown::render("My-Pipeline-Report.Rmd", "pdf_document")
```

4.3 File organization and naming

In many stages of a pipeline, an analyst will want to include details of the settings or what dataset they started out with. Rather than saving a folder or file name that is long and uninformative there are many different ways to organizing folders and files.

Key techniques for this are available and known in the data analysis community as 'Tidy Data' guidelines. In the words of Wickham (2014) the order that data should be arranged in follows some generic principles:

> 'A good ordering makes it easier to scan the raw values. One way of organizing variables is by their role in the analysis: are values fixed by the design of the data collection, or are they measured during the course of the experiment? Fixed variables describe the experimental design and are known in advance. Computer scientists often call fixed variables dimensions, and statisticians usually denote them with subscripts on random variables. Measured variables are what we actually measure in the study. Fixed variables should come first, followed by measured variables, each ordered so that related variables are contiguous. Rows can then be ordered by the first variable, breaking ties with the second and subsequent (fixed) variables.'

4.3.1 An exemplar

The following protocol was developed for an ecology and biodiversity database that the author of this PhD thesis was involved with. The naming convention relied heavily on a sequence of information being used to order the names of folders, subfolders and files. This is:

1. The project name (and optional sub-project name)
2. Data type (such as experimental unit, observational unit, and/or measurement methods)
3. Geographic location (State, Country)
4. Temporal frequency and coverage (such as annual or seasonal tranches).

4.3.2 The concepts of slow moving dimensions and fast moving variables

The concept of dimensions and variables can be useful here, and especially for deciding on filenames. Dimensions are fixed or change slowly while variables change more quickly. By 'change', this means that there are more of them. For example the project name is 'fixed', that is it does not change across the files, but the sub-project name does change, just more slowly (say there may be 2-3 different sub-projects within a project). Then there may be a set of data types, and these 'change' more quickly than the sub-project name. Then the geographic and temporal variables might change quickest of all.

So a general rule for the order of things can be stated. The fixed and slowly changing variables should come first (those things that don't change, or don't change much), followed by the more fluid variables (or things that change more across the project). List elements can then be ordered so that the groups of things that are similar will always be contiguous, and vary sequentially within clusters.

An example is shown in Table 1 to describe this and make it easier to understand. Here is a set of file names that were constructed for an ecological field sites project that I worked on (http://www.supersites.net.au/). That project involved ecological data sampled at plot-based measurement locations. At the beginning of the procedure a controlled vocabulary of data types and their acronyms was created.

Table 1: An example of standardised filename conventions to simplify tracking complicated datasets

Filename	Title
asn_fnqr_soil_charact_robson_2011.csv	Soil Data, Far North Queensland Rainforest SuperSite, Robson Creek, 2011
asn_fnqr_soil_pit_robson_2012.csv	Soil Pit Data, Water Content and Temperature, Far North Queensland Rainforest SuperSite, Robson Creek, 2012
asn_fnqr_veg_seedling_robson_2010-2012.csv	Seedling Survey, Far North Queensland Rainforest SuperSite, Robson Creek, 2010-2012
asn_fnqr_veg_seedling_transect_coord_robson_2010-2012.csv	Seedling Survey, Far North Queensland Rainforest SuperSite, Robson Creek, 2010-2012
asn_fnqr_core_1ha_robson_2014.csv	Soil Pit Data, Soil Characterisation, Far North Queensland Rainforest SuperSite, Robson Creek, Core 1 ha plot, 2014
asn_fnqr_fauna_biodiversity_ctbcc_2012.csv	Vertebrate Fauna Biodiversity Monitoring, Far North Queensland Rainforest SuperSite, CTBCC, 2012
asn_fnqr_fauna_biodiversity_ctbcc_2013.csv	Vertebrate Fauna Biodiversity Monitoring, Far North Queensland Rainforest SuperSite, CTBCC, 2013
asn_fnqr_fauna_biodiversity_ctbcc_capetrib_2014.csv	Avifauna Monitoring, Far North Queensland Rainforest SuperSite, Cape Tribulation, 2014
asn_fnqr_fauna_biodiversity_ctbcc_lu11a_2014.csv	Vertebrate Fauna Biodiversity Monitoring, Far North Queensland Rainforest SuperSite, CTBCC, LU11A, 2014
asn_fnqr_fauna_biodiversity_ctbcc_lu7a_2014.csv	Vertebrate Fauna Biodiversity Monitoring, Far North Queensland Rainforest SuperSite, CTBCC, LU7A, 2014
asn_fnqr_fauna_biodiversity_ctbcc_lu7b_2014.csv	Vertebrate Fauna Biodiversity Monitoring, Far North Queensland Rainforest SuperSite, CTBCC, LU7B, 2014

5 Tracking method steps: Visualisation techniques

5.1 Make a list of steps, inputs and outputs

A very simple example of a pipeline is shown in Table 2. The steps and data listed in Table 2 can be visualised using the `newnode` function described below in case study 3. This creates the graph of this pipeline shown in Figure 5. As the analysis progresses through the phases of testing, refinement and final versions. The linked table and graphical depiction can be very helpful for reference by the analyst. The optional setting to define a status of each step (TODO, DONE, WONTDO) can be used to add colour, and show steps that remain to be done. The addition of short summary descriptions are also very useful for orienting oneself to the required tasks and their priorities. Such flow chart diagrams can be printed up on large sheets of paper and stuck on the wall beside a computer workstation for use in day-to-day work.

Table 2: A table with the steps of a simple data analysis pipeline

STEP	INPUTS	OUTPUTS	DESCRIPTION	STATUS
Step1	Source 1, Source 2	Derived 1, QC check	This might be latitude and longitude of sites	DONE
Step2	Source 3	Derived 2	This might be weather data	DONE
Step3	Derived 1, Derived 2	Derived 3	Merging these data means they can be analysed	TODO
Step4	Derived 3	Model selection		TODO
Step5	Model selection	Sensitivity analysis		TODO

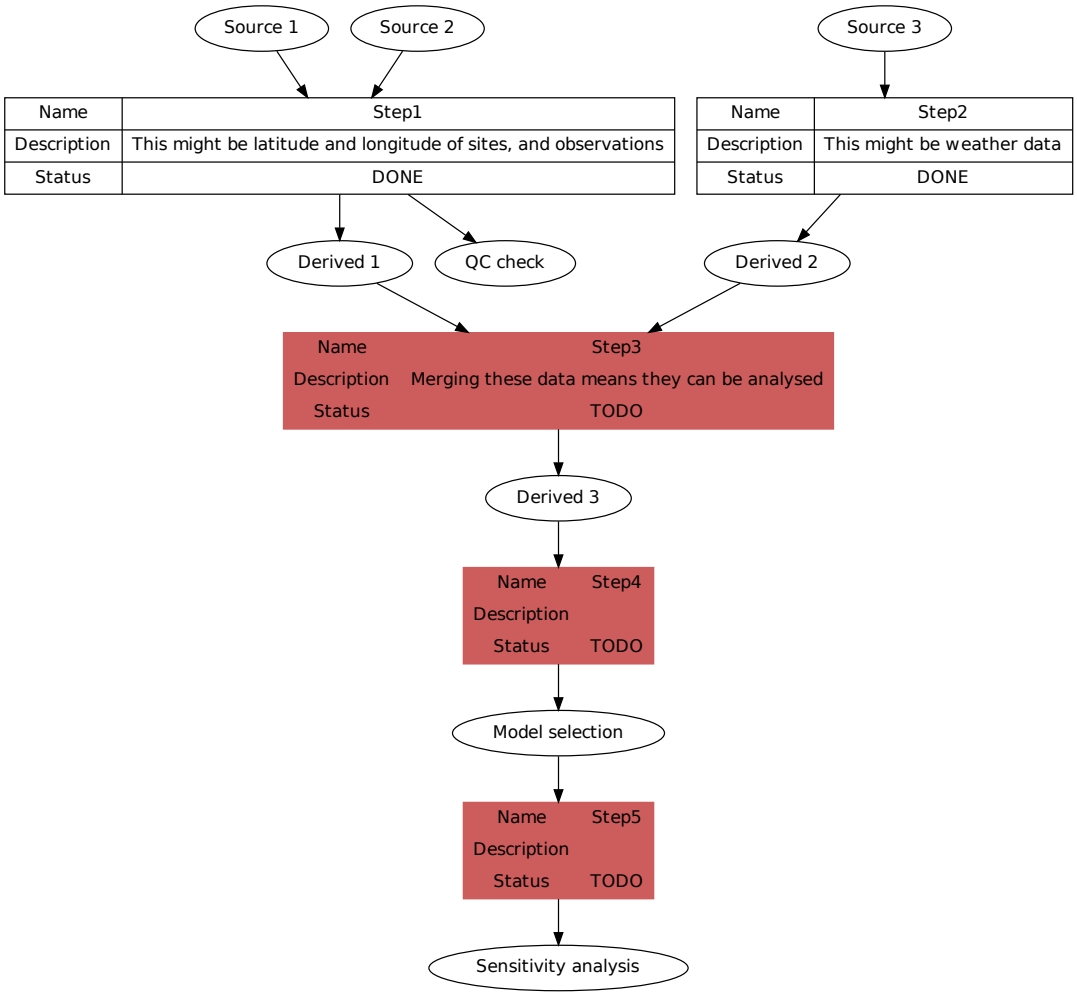


Figure 5: A visualisation of a data analysis pipeline showing the use of colour

As an example of the kinds of tangible steps such a workflow might entail a schematic diagram has been created and is shown in Figure 6.

Figure 6: A schematic flow chart showing the steps required to prepare and conduct an analysis of health, environmental and social data.

A high resolution version of this image is available online at https://github.com/swish-climate-impact-assessment/swish_data_management_procedures/blob/phd_appendix/images/envepi_data_pipeline.pdf.

5.2 Case study 3: Visualisation of methods steps using bespoke software

The method step is the key atomic unit of a scientific pipeline. It consists of inputs, outputs and a rationale for why the step is taken.

A simple way to keep track of the steps, inputs and outputs is shown in Table 3.

The steps and data listed in Table 3 can be visualised. To achieve this an R function was written

Table 3: A simple table to track method steps, data inputs and outputs

STEP	INPUTS	OUTPUTS
Step1	Input 1, Input 2	Output 1
Step2	Input 3	Output 2
Step3	Output 1, Output 2	Output 3

as part of this PhD project and is distributed in the author's own R package available on Github (https://github.com/ivanhanigan/disentangle). This is the `newnode` function. The function returns a string of text written in the `dot` language which can be rendered in R using the `DiagrammeR` package, or the standalone `graphviz` package. This creates the graph view shown in Figure 7. Note that a new field was added for Descriptions as these are highly recommended.

```
library(disentangle); library(stringr); library(readxl)
steps <- read_excel("steps_basic_workflow.xlsx")
nodes <- newnode(indat = steps, names_col = "STEP",
                 in_col = "INPUTS",out_col = "OUTPUTS")
DiagrammeR::grViz(nodes)
```

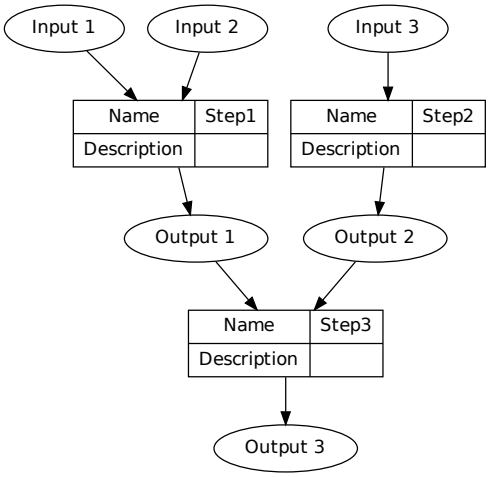


Figure 7: A graphical view of the steps that comprise a simple data analysis pipeline

References for appendix 2

Barnett, A. G. & Dobson A. J. (2010). *Analysing Seasonal Health Data*. Springer, Berlin, Heidelberg, Germany.

Barnett A. G., Baker P., & Dobson A. J. (2014). Package 'season': Analysing seasonal data R functions. R package version 0.3-5.

Bodnar, A., Castorina, R., Desai, M., Duramad, P., Fischer, S., Klepeis, N., Liang S., Mehta S., Naumoff K., Noth E. M., Schei M., Tian L., Vork K. L. & Smith, K. R. (2004). Lessons learned from 'The Skeptical Environmentalist': an environmental health perspective. *International Journal Of Hygiene And Environmental Health*, 207(1):57–67.

Borer E. T., Seabloom E. W., Jones M. B. & Schildhauer M. (2009). Some simple guidelines for effective data management. *Bulletin of the Ecological Society of America*, 205–214.

Cai W., Cowan T., Braganza K., Jones D. & Risbey J. (2010). Comment on 'On the recent warming in the Murray Darling Basin: Land surface interactions misunderstood' by Lockart et al. *Geophysical Research Letters*, 37(10):1–3.

Fleming L., Haines A., Golding B., Kessel A., Cichowska A., Sabel C., Depledge M., Sarran C., Osborne N., Whitmore C., Cocksedge N., & Bloomfield D. (2014). Data mashups: Potential contribution to decision support on climate change and health. *International Journal of Environmental Research and Public Health*, 11(2):1725–1746, 2014.

Gentleman R. & Temple Lang D. (2004). Statistical Analyses and Reproducible Research. *Journal of Computational and Graphical Statistics*, 16(1):1–23.

Healy K. J. (2013). Choosing Your Workflow Applications. URL http://kieranhealy.org/files/papers/workflow-apps.pdf [Accessed 20 Aug. 2016].

Leek J. T. & Peng R. D. (2015a). Opinion: Reproducible research can still be wrong: Adopting a prevention approach. *Proceedings of the National Academy of Sciences of the United States of America*, 112(6):1645–1646.

Leek J. T. & Peng R. D. (2015b). Statistics: P values are just the tip of the iceberg. *Nature*, 520 (7549):612–612.

Long, J. S. (2008). *The Workflow Of Data Analysis Using Stata*. Stata publishing, 2008. ISBN 9781597180474.

McCullough B. D. & Heiser D. A. (2008). On the accuracy of statistical procedures in Microsoft Excel 2007. *Computational Statistics and Data Analysis*, 52(10):4570–4578.

McMichael A. J. (2013). Impediments to comprehensive research on climate change and health. *International Journal of Environmental Research and Public Health*, 10(11):6096–6105.

Michener W. K., Brunt J. W., Helly J. J., Kirchner T. B. & Stafford S. G. (1997). Nongeospatial metadata for the ecological sciences. *Ecological Applications*, 7(1):330–342.

Noble W. S. (2009). A quick guide to organizing computational biology projects. *PLoS Computational Biology*, 5(7):1–5.

O'Keefe. B. (2007). Hackers pick up UQ cash prize, 2007. URL http://www.theaustralian.com.au/higher-education/hackers-pick-up-uq-cash-prize/story-e6frgcjx-1111113191659 [Accessed 14 Oct. 2015].

Peng R. D. (2011). Reproducible research in computational science. *Science*, 334(6060):1226–1227.

Peng R. D. (2015). *Report Writing For Data Science In R.* Leanpub. Unpublished Draft. URL https://leanpub.com/reportwriting [Accessed 22 Dec. 2015].

Peng R. D. & Dominici F. (2008). *Statistical Methods For Environmental Epidemiology With R. A Case Study In Air Pollution And Health.* Springer Science & Business Media, New York, USA.

Peng R. D., Dominici F., Zeger S. L. (2006). Reproducible epidemiologic research. *American Journal of Epidemiology*, 163(9):783–789.

Pullin A. S. & Salafsky N. (2010). Save the whales? Save the rainforest? Save the data! *Conservation Biology*, 24(4):915–917.

Schulte E., Davison D., Dye T., & Dominik C. (2012). A multi-language computing environment for literate programming and reproducible research. *Journal of Statistical Software*, 46(3).

Schwab M., Karrenbach M., & Claerbout J. (2000). Making scientific computations reproducible. *Computing in Science and Engineering*, 2(6):61–67.

Scott T. A. (2010). Reproducible Research with R, TEX and Sweave, 2010. URL http://biostat.mc.vanderbilt.edu/wiki/pub/Main/TheresaScott/ReproducibleResearch.TAScott.handout.pdf [Accessed on 30 Jan. 2016].

Silberzahn R. & Uhlmann E. L. (2015). Crowdsourced research: Many hands make tight work. *Nature*, 526(7572):189–191.

Sólymos P. & Fehér Z. (2008). The mefa package: a tool for reproducible data processing in biogeography. URL http://biogeography.blogspot.com.au/2008/04/mefa-package-tool-for-reproducible-data.html [Accessed on 10 Aug. 2015].

Vines T. H., Albert A. Y. K., Andrew R. L., Débarre F., Bock D. G., Franklin M. T., Gilbert K. J., Moore J. S., Renaut S., & Rennison D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, 24(1):94–97.

White E. P., Baldridge E., Brym Z. T., Locey K. J., Mcglinn D. J., & Supp S. R. (2013). Nine simple ways to make it easier to (re)use your data. *Ideas in Ecology and Evolution*, 6(2):1–10

Wickham H. (2014). Tidy data. *Journal of Statistical Software*, 59(10):1–23.

Yihui X. (2014). Chapter 1. Knitr: A comprehensive tool for reproducible research in R. In Stodden V., Leisch F., & Peng R. D., editors, *Implementing Reproducible Research.* CRC Press, Boca Raton, USA. ISBN 978-0-387-98140-6.

Appendix 3: Decisions to be made during statistical modelling

This section focuses on decision making during statistical modelling. The case of investigating effect modification is used as an example. The analyst has several options available to them when constructing the model parametrisation for adjustment to explain modification effectively. Unlike in confounding when it is useful to assess if there were a substantial change of the effect estimate when controlling for a third variable, models including effect modification can be hard to interpret. Taking account of effect modification becomes increasingly important when modelling complex systems.

An example is provided by paper 10 of this thesis (Vally *et al.* 2012), in which the relationship between proximity to wetlands and Ross River virus infection may be different for towns and rural areas, where the 'urban' variable is the modifier.

There are three common ways for data analysts to address this question in statistical models:

1. separate models for each group
2. interaction terms
3. re-parametrisation to explicitly depict interactions.

The first way is to split apart the dataset and conduct separate analyses of multiple groups. For example, one could run the regression in the urban zone and the rural zone separately and see if there were different exposure response functions in the two models. This was the approach of paper 10 of this thesis and Figure 1 below is reproduced from the results of that paper to illustrate this discussion. This is an example of a data analysis with a very large number of alternatives that could have been implemented, and the resulting report of the particular analysis that was done can be investigated in more depth by accessing the analytic dataset and code from the authors website (Hanigan 2015).

The approach to subset the data into two sets and compute separate models has the strength that it is simple to do and yields results that are easy to interpret. A limitation of this method is that by splitting the dataset one loses degrees of freedom and therefore statistical power. It is also not possible to use model selection procedures like assessing the change in Akaike's Information Criterion (AIC) for model comparisons because observed data are different between the datasets used to fit the models.

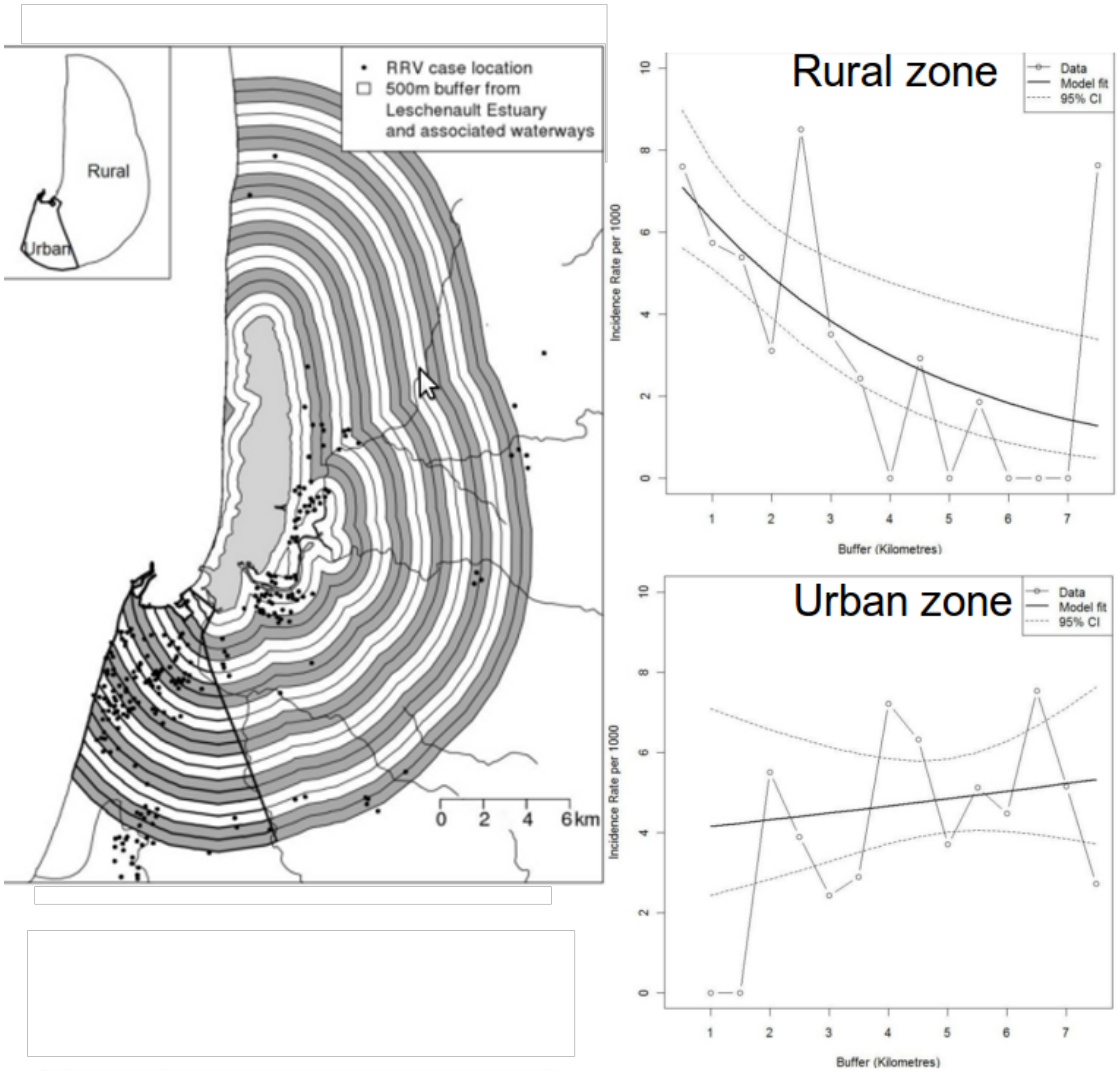


Figure 1: Results from paper 10 of this thesis: Vally et al. (2012) showing the change in risk of Ross River virus and proximity to waterways

The second approach (which removes limitations of the first option) is to fit interaction terms such as the multiplicative interaction model described by Brambor et al. (2006). The statistical method can be easily implemented in software by including a multiplicative term between two variables, however in practice the resulting post-estimation regression outputs can be difficult to interpret. For example, say one wants to calculate the effect and standard error for exposure X on health outcome Y with the interaction of the effect modifier Z. The form of this model can be written as:

$$Y = \beta_1 X + \beta_2 Z + \beta_3 XZ$$

where β_1, β_2 and β_3 are the regression coefficients estimated and the term XZ is the interaction between exposure and effect modifier. In the case of paper 10 of this thesis the proximity to waterways is variable X, while Z would be coded as 1 in the urban zone and 0 otherwise. Thus the AIC of this model is directly comparable to the null model without this interaction term and a judgement of the improvement in model performance can be made.

The weakness of this approach however is the difficulty for interpretation that comes when using this method for calculating the marginal effect of X on Y and the conditional standard error. The specific method described in Brambor et al. is:

1. Calculate the coefficients and the variance-covariance matrix from the regression model.
2. β_1 is the main effect (that is the association between proximity and RRv in the rural zone) while the marginal effect is $\beta_1 + \beta_3$ and represents the effect when condition Z is 1 (the proximity to wetlands experienced inside the urban zone).
3. The conditional standard error is:

$$\sqrt{var(\beta_1) + Z^2 var(\beta_3) + 2Zcov(\beta_1 \beta_3)}$$

Therefore the strengths of this approach are that it does not reduce degrees of freedom and is straightforward to specify the model in standard statistical software packages. The limitations are related to interpretation of the resulting coefficients for both main and marginal effects, and standard errors for these.

The third approach available to analysts makes it easier to interpret and use the resulting regression output. This method was employed in paper 1 in the final modelling phase in which estimates were calculated for the different drought exposure-response functions in each of the sub-groups. In the terms of Brambor et al. (2006) it is simple to:

1. Calculate X1 = X * Z (i.e = exposure condition is met, zero otherwise).

2. Calculate X0 = X * (1-Z) (i.e. = exposure for NON-condition, zero otherwise).
3. Instead of X, Z and XZ, fit X1, X0 and Z. This model also contains three parameters and captures the same interactions as it is the same model with a different parametrisation. The standard errors for the X1 and X0 coefficients are available directly from the regression output.

This method is much easier to interpret and is considerably more flexible than the other two approaches when many interaction terms are being compared. A limitation remains for this method in that the pre-processing steps required are more complicated, and there are inherently more possibilities for the data analyst to make errors in writing their code as they make these changes to the analytical data. If each of these steps are executed with reproducible code, any errors in the execution will be easier to find and rectify.

References for appendix 3

Brambor, T., Clark, W.R. & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, 14(1), 63–82.

Hanigan, I.C. (2015). Reproducible code for Vally 2012 Ross River Rates. https://github.com/ivanhanigan/Vally2012-RossRiverRates [Accessed 26 Jan. 2016]

Vally, H., Peel, M., Dowse, G.K., Cameron, S., Codde, J.P., Hanigan, I. & Lindsay, M.D. a. (2012). Geographic information systems used to describe the link between the risk of Ross River virus infection and proximity to the Leschenault Estuary, WA. *Australian and New Zealand Journal of Public Health*, 36(3), 229–235.