Climate and Environment as Signal Predictors of Ross River Virus Disease

Rosalie Ellen Woodruff

A thesis submitted for the degree of Doctor of Philosophy
National Centre for Epidemiology and Population Health
The Australian National University, May 2003
Declaration

This thesis is my original work, except where indicated.

Rosalie Ellen Woodruff
CHAPTER 1: RESEARCH CONTEXT .................................................. 1

1.1 INTRODUCTION ..................................................................... 1

1.2 THE BURDEN AND CONTROL OF MOSQUITO-BORNE DISEASES .................................................................................. 2

1.2.1 Main mosquito-borne diseases - global level ........................................ 2

1.2.2 Main mosquito-borne diseases - Australia ........................................... 5

1.2.2.1 Ross River virus disease ........................................ 5

1.2.2.2 Murray Valley encephalitis ........................................... 6

1.2.2.3 Barmah Forest virus .................................................... 7

1.2.2.4 Malaria ........................................................................... 7

1.2.2.5 Dengue ........................................................................... 8

1.2.2.6 Japanese encephalitis .................................................... 9

1.2.3 Determinants of mosquito-borne disease transmission .......................... 9

1.2.4 Climate as a tool for early warning of mosquito-borne diseases ............ 10

1.2.5 Rationale for focusing on Ross River virus disease ............................... 11

1.3 ECOLOGY OF ROSS RIVER VIRUS ........................................ 12

1.3.1 Ross River virus .............................................................. 12

1.3.1.1 The RRv transmission cycle ......................................... 12

1.3.1.2 Endemic and epidemic cycles of transmission .......................... 13

1.3.1.3 Virus persistence and re-introduction mechanisms .......................... 14

1.3.1.4 Geographic distribution of the virus .................................. 15
1.3.2 Vector species

1.3.2.1 *Culex annulirostris* ..................................................... 16
1.3.2.2 *Aedes vigilax* and *Aedes camptorynchus* ...................... 18
1.3.2.3 Other possible RRv vectors ........................................ 19

1.3.3 Reservoir hosts ............................................................. 21

1.3.3.1 Evidence implicating vertebrate hosts ......................... 21
1.3.3.2 Distribution and breeding cycle of the main reservoir host species ........................................ 22
1.3.3.3 Vertebrate immunity .................................................. 25

1.4 Ross River Virus in Humans .............................................. 27

1.4.1 Early understandings of the disease .................................. 27
1.4.2 Illness and incapacity .................................................. 30
1.4.2.1 Symptoms ................................................................. 30
1.4.2.2 Duration ................................................................. 31
1.4.2.3 Treatment and prevention ......................................... 31
1.4.3 Incubation period ......................................................... 32
1.4.4 Incidence and prevalence of RRv infection ......................... 33
1.4.5 Geographic and temporal distribution ............................... 34
1.4.5.1 Within Australia ..................................................... 34
1.4.5.2 Outside Australia .................................................... 35
1.4.5.3 A rural disease? ....................................................... 35
1.4.6 Factors moderating RRv transmission and infection ............... 36
1.4.6.1 Age, sex, and race .................................................. 36
1.4.6.2 Human behaviour and lifestyle ................................... 37
1.4.6.3 Land use changes .................................................... 37

1.5 Weather, Climatic and Environmental Influences on the Arboviral Cycle .... 38

1.5.1 The weather-climate distinction ..................................... 38
1.5.2 Rainfall ................................................................. 39
1.5.3 Temperature and humidity ............................................. 42
1.5.4 The El Niño-Southern Oscillation ................................ 44
1.5.4.1 Description ............................................................ 44
1.5.4.2 ENSO and mosquito-borne diseases: global teleconnections .... 46
1.5.4.3 ENSO teleconnections in Australia .................................... 47
1.5.4.4 The use of ENSO as an early warning indicator .................. 48
1.5.5 Tide height and sea-level ............................................. 52
1.5.6 Environment ............................................................. 52
1.5.6.1 Vegetation ............................................................. 52
1.5.6.2 The complex role of irrigation ................................... 53
1.5.6.3 Mosquito control .................................................... 54
1.5.7 The influence of climate change on mosquito-borne diseases ........... 57
1.5.7.1 Global changes to climate ......................................... 57
CHAPTER 2: RESEARCH QUESTIONS AND THESIS SYNOPSIS .. 67

2.1 RESEARCH QUESTIONS .................................................... 67
2.2 THESIS SYNOPSIS ........................................................... 68

CHAPTER 3: DATA COLLECTION ............................................. 71

3.1 INTRODUCTION ............................................................... 71
3.2 DISEASE DATA ............................................................... 71
  3.2.1 Type and structure ....................................................... 71
  3.2.2 Disease data quality .................................................... 73
    3.2.2.1 Inconsistencies with the case definition .................. 73
    3.2.2.2 Reporting quality .................................................. 75
3.3 DEMOGRAPHIC DATA ....................................................... 77
3.4 CLIMATE AND ENVIRONMENT DATA .............................. 77
  3.4.1 Climatic indices ......................................................... 78
    3.4.1.1 Southern Oscillation Index ....................................... 78
    3.4.1.2 Sea surface temperature ........................................ 79
  3.4.2 Weather data ............................................................. 79
    3.4.2.1 Interpolated weather surface data .......................... 80
    3.4.2.2 Quality of the interpolated surface ........................ 83
    3.4.2.3 Long-term average climate data ............................. 84
  3.4.3 Tidal data ................................................................. 84
  3.4.4 Mosquito data ........................................................... 85
  3.4.5 Mosquito control activities ......................................... 86
    3.4.5.1 Murray River ....................................................... 86
    3.4.5.2 Southwest .......................................................... 86
  3.4.6 Irrigation data ........................................................... 87

CHAPTER 4: EPIDEMIOLOGY OF RRV DISEASE IN AUSTRALIA .. 89
  1991-1999

4.1 INTRODUCTION ............................................................. 89
4.2 METHODS ................................................................. 90
  4.2.1 Spatial analysis .......................................................... 90
  4.2.2 The base spatial unit ................................................... 91
  4.2.3 Georeferencing the disease data ................................... 92
### Chapter 4: Results

- **4.2.4** Calculating population rates ................................................................. 95
- **4.2.5** Calculating expected cases ................................................................. 96
- **4.2.6** Rainfall zones ....................................................................................... 99

#### Results

- **4.3.1** Notification rates .................................................................................. 100
- **4.3.2** Age and sex distribution ......................................................................... 101
- **4.3.3** Reporting period .................................................................................... 105
- **4.3.4** Spatial distribution
  - **4.3.4.1** By rainfall zones ............................................................................. 106
  - **4.3.4.2** By SLA ......................................................................................... 106
- **4.3.5** Seasonal effects ................................................................................... 108

#### Discussion

- **4.4.1** Data quality ......................................................................................... 116
- **4.4.2** Sex differences in notifications ......................................................... 118
- **4.4.3** Patterns of RRv distribution in Australia .......................................... 119
- **4.4.4** An increase in cases of RRv disease? .............................................. 120

#### Conclusions

- **4.5** Conclusions .............................................................................................. 121

---

### Chapter 5: Case Study Methods

- **5.1** Introduction .............................................................................................. 123
- **5.2** Developing an RRv Geographical Classification .................................... 124
  - **5.2.1** Scale of analysis .............................................................................. 124
  - **5.2.2** Existing regional classification systems ...................................... 126
  - **5.2.3** Steps in building the RRv regional system
    - **5.2.3.1** Step 1 - Long-term average climate maps ............................... 127
    - **5.2.3.2** Step 2 - Assigning SLAs to climate zone attributes ............. 128
    - **5.2.3.3** Step 3 - Forming the RRv bioclimatic regions .................... 129
  - **5.2.4** Qualifications on the method ............................................................. 145
- **5.3** Outcome Variable .................................................................................. 146
  - **5.3.1** Epidemics or incidence? ................................................................. 146
  - **5.3.2** Definition of an epidemic ................................................................. 147
- **5.4** Predictive Modelling ............................................................................... 149
  - **5.4.1** Introduction ..................................................................................... 149
  - **5.4.2** Criteria for selecting the study areas
    - **5.4.2.1** Important criteria .................................................................... 150
    - **5.4.2.2** Desirable criteria .................................................................... 152
  - **5.4.3** The two case study areas
    - **5.4.3.1** Murray River study area ......................................................... 153
    - **5.4.3.2** Southwest study area .............................................................. 153
  - **5.4.4** SLA exclusion criteria ...................................................................... 154
5.4.5 The modelling method ............................................................................................... 155
5.4.5.1 Step 1 - Single variable analysis ........................................................................ 156
5.4.5.2 Steps 2 and 3 - Early and late warning models .............................................. 158
5.4.5.3 Step 4 - Criteria for best-fitting models ......................................................... 159
5.4.5.4 Step 5 - Cross-validation .................................................................................. 160
5.4.5.5 Combined region analysis in the Murray study area ...................................... 161
5.4.5.6 Managing dependency in the data ................................................................. 162

5.4.6 Confounding and interaction .................................................................................. 162
5.4.6.1 Mosquito control activities .............................................................................. 162
5.4.6.2 Irrigated regions .............................................................................................. 163
5.4.6.3 Human immunity .............................................................................................. 163

CHAPTER 6: THE MURRAY RIVER STUDY AREA .................................................. 165

USING CLIMATE DATA TO PREDICT EPIDEMICS

6.1 ABSTRACT .............................................................................................................. 165
6.2 PROFILE OF THE STUDY AREA .......................................................................... 166
6.2.1 Population and land use ...................................................................................... 166
6.2.2 Climate and environment .................................................................................... 173
6.3 ECOLOGY OF RRV IN THE MURRAY .................................................................... 174
6.3.1 Vectors .................................................................................................................. 174
6.3.2 Vertebrate hosts ................................................................................................... 179
6.4 EPIDEMIOLOGY OF RRV DISEASE ...................................................................... 180
6.4.1 Historical report of outbreaks ............................................................................. 180
6.4.2 Study period ......................................................................................................... 181
6.5 HYPOTHESES ......................................................................................................... 187
6.6 METHODS ................................................................................................................. 189
6.7 RESULTS .................................................................................................................. 189
6.7.1 Single variable analysis ...................................................................................... 189
6.7.1.1 Effect of rainfall ............................................................................................... 190
6.7.1.2 Effect of the SOI and SST ................................................................................ 205
6.7.1.3 Effect of temperature ...................................................................................... 205
6.7.1.4 Effect of humidity ........................................................................................... 206
6.7.2 Early warning models (months 1-5) ................................................................... 207
6.7.2.1 Region 1 early warning model ......................................................................... 207
6.7.2.2 Region 2 early warning model ......................................................................... 210
6.7.3 Late warning models (months 1-8) ..................................................................... 212
6.7.3.1 Region 1 late warning model .......................................................................... 212
6.7.3.2 Region 2 late warning model .......................................................................... 215
6.7.4 Combined region models

6.7.4.1 Combined regions early warning model

6.7.4.2 Combined regions late warning model

6.7.5 Summary

6.8 Discussion

6.8.1 Limitations of the study

6.8.2 Host immunity

6.8.3 The influence of rainfall

6.8.3.1 The role of irrigation

6.8.4 The influence of temperature

6.8.5 Analysis within RRv bioclimatic regions

6.8.6 Precision and validity of the models

6.8.7 Implications for climate change

6.8.8 Generalisability of results

CHAPTER 7: THE SOUTHWEST STUDY AREA

USING CLIMATE AND MOSQUITO DATA TO PREDICT EPIDEMICS

7.1 Abstract

7.2 Profile of the study area

7.2.1 Population and land use

7.2.2 Climate and environment

7.3 Ecology of RRv in the Southwest

7.3.1 Vectors

7.3.2 Vertebrate hosts

7.4 Epidemiology of RRv Disease

7.4.1 Historical

7.4.2 Study period

7.5 Analytic methods particular to the Southwest

7.5.1 Additional outcome variable tested

7.5.2 Including the mosquito data in the modelling

7.5.3 Accounting for the influence of tidal height

7.5.4 Reduced data for the late warning model

7.6 Hypotheses

7.7 Results

7.7.1 Single variable analysis

7.7.1.1 Effect of rainfall

7.7.1.2 Effect of the SOI and SST

7.7.1.3 Effect of temperature
8.6 THE CONTEXT FOR RRv EARLY WARNING IN AUSTRALIA ................................................. 300
8.6.1 Current RRv disease management ......................................................................... 300
  8.6.1.1 Surveillance .................................................................................................. 301
  8.6.1.2 Vector control ............................................................................................. 302
  8.6.1.3 Public notification and education .................................................................. 303
8.6.2 How could the current approach be improved? ..................................................... 304
  8.6.2.1 Ineffective prevention education .................................................................. 304
  8.6.2.2 Economic, environmental, and social sustainability ...................................... 305
  8.6.2.3 Tourism in Australia .................................................................................... 307
  8.6.2.4 Population pressures .................................................................................... 307
8.6.3 Summary .................................................................................................................. 308

8.7 A HYPOTHETICAL RRv EARLY WARNING SYSTEM .................................................... 308
8.7.1 Overview of the early warning system .................................................................. 308
  8.7.1.1 Stage One (epidemic forecasting) .................................................................. 309
  8.7.1.2 Stage Two (early warning) .......................................................................... 309
  Stage Three .............................................................................................................. 310
  8.7.1.4 Inputs required ............................................................................................. 310
8.7.2 General constraints on early warning systems ....................................................... 313
8.7.3 The benefits of early warning ................................................................................ 314

8.8 CONCLUSIONS ........................................................................................................... 315

CHAPTER 9: CONCLUSIONS AND FURTHER RESEARCH ............................................. 317

APPENDIX A .................................................................................................................. 323
AWARD, PUBLICATIONS, AND CONFERENCE PRESENTATIONS .................................. 323
A.1 Award ....................................................................................................................... 323
A.2 Publications ............................................................................................................. 323
A.3 Relevant presentations at conferences ..................................................................... 324

REFERENCES ................................................................................................................ 325
Acknowledgements

Supervisory Panel

Dr Charles Guest (NCEPH, ANU) was the chair of the Supervisory Panel. Charles provided the initial direction for this project, and spent many hours helping me establish the necessary interdisciplinary links. In addition to providing unqualified support on such weighty matters as epidemiological principles and methods, he has also guided me through the etiquette of academic research and practice. His interest (beyond the call of duty) in my writing and language has provided us with countless hours of amusement. His warmth and friendliness never interfered with his ability to provide critical feedback on my work. Overall, I want to thank him for the extraordinary balance of encouragement and a continual pressure to maintain deadlines that he managed to sustain throughout.

Dr Graeme Garner is a Principal Scientist with the Animal Health Science Unit (Department of Agriculture, Fisheries, and Forestry Australia). Without his expertise and support, I doubt I would have embarked on an epidemiological thesis that required a thorough understanding of Ross River viral transmission cycles. His methodical and practical approach to scientific research has been a great influence. He introduced me to GIS methodology, and helped provide the means for my continued access to that software throughout the period of the thesis. Most importantly, his cheery disposition and confidence in my ability to complete the task have been a very valuable contribution.

Prof. Niels Becker (NCEPH, ANU) provided me with statistical direction and support, and I am extremely grateful for his help and forbearance, given my lack of knowledge in that area. Dr Janette Lindesay (Geography, ANU) was the formal meteorological link in this work, and I thank her for her support. She was the first person I met who described
the Australian climate with excitement and passion, and her enthusiasm has been infectious.

**Funding, data provision, and advice**

This work was made possible with the financial support of an Australian Post-Graduate Award and a supplementary NCEPH scholarship. Additional funding was also provided by the Electric Power Research Institute, Palo Alto, California (a portion of the $118,000 grant covered the purchase of data and additional salary expenses). Dr Kristie Ebi from that organisation provided excellent critical comment on several parts of this work that have been prepared for publication.

I would like to thank staff from the Animal Health Science Unit of the Commonwealth Department of Agriculture, Fisheries, and Forestry Australia, who provided me with a desk, computing resources, GIS software, and library access throughout the period of study. In addition to Graeme Garner - Geoff Gard, Mike Nunn, Rob Cannon, Terry Nicholls, David Adams, and Peter Black provided collegial support, good humour, and advice about a scientific approach to research.

Terry Carvan (Greater Murray Health Centre, NSW Department of Health) was an invaluable source of local knowledge regarding RRv transmission in the Murray study area. He also helped to clarify the realities of competing health priorities in rural public health, which ultimately influenced the study design and my thinking about the requirements for early warning systems for Ross River virus disease. Dr Michael Lindsay (Microbiology Department, University of Western Australia) kindly made himself available to discuss many aspects of Ross River virus ecology, and also helped me to assemble the mosquito trapping data for the Southwest study area.

Keith Moodie (Silo, Queensland Department of Natural Resources) spent many hours discussing the pros and cons of different methods for preparing the weather data, and I am thankful for his patience. The National Notifiable Diseases Surveillance System, Communicable Diseases Network of Australia, provided the Ross River virus disease notifications. The Murray-Darling Basin Commission provided data on releases from the Hume Reservoir.
Family and friends

My partner, Paul Gibson, has assisted me with constant support over the years. He has helped me remember that a PhD is but one element of a well-rounded life. Consequently, I have enjoyed the work involved from the beginning to the end. He also helped enormously with structural and detailed comments on the text, and tolerated my mood swings throughout two "PhD pregnancies". Our daughter, Mae, has been particularly understanding about my absorption with "data" for the "PhE".

My parents, Patricia and Brian Woodruff, provided me with the confidence and resolve I needed to complete this task. All my sisters have been supportive, and Alice Woodruff in particular provided helpful editing advice. Several friends and fellow PhD students have been generous with technical help or morale-boosting support. Many thanks to Jenny Gibson, Keith Harrison, Lisa Alleva, Chris Kelman, Su Wildriver, Cathy Banwell, Eileen Wilson, Jacqui de Chazal, Colin Butler, Susan Nancarrow, and Anne Gardner. I would also like to acknowledge the support of academic and administrative staff at NCEPH over the years.
In this thesis, reference to other work is indicated in the following way:

- Where a table or figure has been reproduced without change, this is credited at the bottom with the word “source”.
- Where a table or figure has been compiled from several published sources, the words “compiled from sources” are used to indicate this.
- Where I have developed a table from published text, the words “adapted from” indicate this.

To save repetition in the tables and figures, I have avoided referencing several of the most common sources from which data were obtained. Unless otherwise noted:

- Ross River virus disease data for the 1991 to 1999 period were provided by the National Notifiable Diseases Surveillance System, Communicable Diseases Network Australia.
- Australian population data (1996) and the Australian Standard Geographical Classification System boundaries were provided by the Australian Bureau of Statistics (specifically, the Integrated Regional Database).
- The Australian climate zones (rainfall, temperature and humidity, and vegetation) were provided by the Australian Bureau of Meteorology.

During the latter course of this PhD, a taxonomic revision of the genus *Aedes* and its subgenera occurred, with the result that several subgenera have been raised to generic rank and have changed name from *Aedes* to *Ochlerotatus* (NSW Arbovirus Surveillance & Vector Monitoring Program 2003a). For ease of citing references, and to avoid confusion with previous published work, I have retained the use of the older classification system throughout this thesis (i.e., *Aedes*).
List of figures and tables

Tables

Table 1.1  Main vector-borne diseases, populations at risk and affected, and current assumed distribution.

Table 1.2  Proposed mechanisms of survival for Australian arboviruses.

Table 1.3  Major mosquito species implicated as vectors of Ross River virus.

Table 1.4  Summary of estimated incubation period for RRv infection.

Table 1.5  The effect of temperature on breeding time and extrinsic incubation periods for the major Ross River virus species.

Table 3.1  Structure of the RRv disease data set.

Table 3.2  Main climatic and environmental factors influencing the life-cycle of RRv, its vectors and vertebrate hosts.

Table 3.3  Classification of La Niña (cold = C) and El Niño (warm = W) episodes by season for the period of the RRv disease study.

Table 4.1  Australian Standard Geographical Classification (ASGC) 1996.

Table 4.2  The major Australian rainfall zones.

Table 4.3  RRv disease notifications and rates per 100 000 per annum, and average annual notification rates, by State and Territory of notification (July 1991 to June 1999).

Table 4.4  RRv disease notifications, median age of notification, and median age of the general population, by State and Territory (July 1991 to June 1999).

Table 4.5  National RRv disease notifications for males and females, sex ratio, general population ratio, and sex status unknown, by State and Territory (July 1991 to June 1999).

Table 4.6  Time from estimated onset of disease symptoms to the report of an RRv disease notification (percentage), by State and Territory (July 1991 to June 1999).

Table 4.7  RRv disease notification rates per 100 000 per annum, by main Australian rainfall zones.

Table 4.8  Percentage of RRv infections each season, by rainfall zone.
Table 5.1  Description of the temperature and humidity zone classification.
Table 5.2  Definitions of 'epidemic' used in related mosquito-borne disease studies.
Table 5.3  Summary of potential RRv disease study areas against the selection criteria.
Table 5.4  Definitional terms relating to the study design for this thesis.
Table 5.5  Climate and environmental variables tested in the predictive models.
Table 6.1  Population characteristics of the two Murray study regions.
Table 6.2  Breakdown of mosquito type from trapping conducted along the Murray River in Victoria, 1991/95.
Table 6.3  Temperature influence on the larval development, population growth, and survival of Culex annulirostris.
Table 6.4  The seasonal pattern of historic RRv disease epidemics and rainfall in the Murray Valley area, summarised from the literature.
Table 6.5  Number of SLAs experiencing epidemics in the two Murray study regions, by year.
Table 6.6  Odds ratios (OR), 95% confidence intervals (CI), coefficients, and standard errors (SE) for the early warning logistic regression model for Murray Region 1.
Table 6.7  Sensitivity, specificity, positive predictive values, and negative predictive values for the early warning model in Murray Region 1, by year.
Table 6.8  Accuracy of the early warning model for Murray Region 1, using varying cut-off values to differentiate the probability of an epidemic, by year.
Table 6.9  Odds ratios (OR), 95% confidence intervals (CI), coefficients, and standard errors (SE) for the early warning logistic regression model for Region 2.
Table 6.10 Sensitivity, specificity, positive predictive values, and negative predictive values for the early warning model in Murray Region 2, by year.
Table 6.11 Accuracy of the early warning model for Murray Region 2, using varying cut-off values to differentiate the probability of an epidemic, by year.
Table 6.12 Odds ratios (OR), 95% confidence intervals (CI), coefficients, and standard errors (SE) for the late warning logistic regression model in Region 1.
Table 6.13 Sensitivity, specificity, positive predictive values, and negative predictive values for the late warning model in Murray Region 1, by year.
Table 6.14 Accuracy of the late warning model for Murray Region 1, using varying cut-off values to differentiate the probability of an epidemic, by year.
Table 6.15 Odds ratios (OR), 95% confidence intervals (CI), coefficients, and standard errors (SE) for the late warning logistic regression model for Region 2.
Table 6.16 Sensitivity, specificity, positive predictive values, and negative predictive values for the late warning model in Murray Region 2, by year.
Table 6.17 Accuracy of the late warning model for Murray Region 2, using varying cut-off values to differentiate the probability of an epidemic, by year.
Table 6.18  Odds ratios (OR), standard errors (SE), and 95% confidence intervals (CI) for the combined regions early warning logistic regression model.

Table 6.19  Sensitivity, specificity, positive predictive values, and negative predictive values for the combined regions early warning model, all years.

Table 6.20  Odds ratios (OR), standard errors (SE), and 95% confidence intervals (CI) for the combined regions late warning logistic regression model. Region interactions added where required.

Table 6.21  Sensitivity, specificity, positive predictive values and negative predictive values for the combined regions late warning model, all years.

Table 6.22  Accuracy of the early warning and late warning models for the two Murray study regions, by year.

Table 6.23  Sensitivity and specificity of the early warning and late warning models for the two Murray study regions, all years.

Table 7.1  Characteristics of the Southwest study area.

Table 7.2  Breakdown of mosquito type from trapping conducted in the Mandurah to Bunbury regions of the Southwest study area, 1988/89.

Table 7.3  The seasonal pattern of historic RRv disease epidemics and rainfall and tides in the Southwest study area, summarised from the literature.

Table 7.4  Number of epidemic Statistical Local Areas in the Southwest study area by year, using two different epidemic definitions.

Table 7.5  Single variable logistic regression model results for mosquito density in the Southwest study area: odds ratios (OR), standard errors (SE), and confidence intervals (CI).

Table 7.6  Southwest early warning logistic regression model: odds ratios (OR), 95% confidence intervals (CI), coefficients, and standard errors (SE).

Table 7.7  Sensitivity, specificity, positive predictive values, and negative predictive values for the early warning model in the Southwest.

Table 7.8  Southwest sub-region early warning logistic regression model, with mosquito density data: odds ratios (OR), 95% confidence intervals (CI), coefficients, and standard errors (SE).

Table 7.9  Sensitivity, specificity, positive predictive values, and negative predictive values for the early warning model in the Southwest sub-region – with mosquito data.

Table 7.10  Southwest sub-region late warning logistic regression model, with mosquito density data: odds ratios (OR), 95% confidence intervals (CI), coefficients, and standard errors (SE).

Table 7.11  Sensitivity, specificity, positive predictive value, and negative predictive value for the early warning model in the Southwest sub-region – with mosquito data.

Table 7.12  Summary table: predictor variables for the early warning and late warning models for the Southwest study area.
Table 7.13
Summary table: sensitivity, specificity, positive predictive values, negative predictive values, and accuracy of the early warning and late warning models for the Southwest study area, and the Southwest sub-region with mosquito data.

Table 8.1
Partial estimates of the annual cost of RRv disease in Australia, for the period 1991-2000.

FIGURES

Figure 1.1
A simplified representation of the distribution of the three major kangaroo populations in Australia.

Figure 1.2
Areas in Australia most likely to receive less than normal rainfall during an El Niño.

Figure 1.3
Areas in Australia most likely to receive more than normal rainfall during a La Niña.

Figure 1.4
Representation of the determinants of Ross River virus disease.

Figure 4.1
Main Australian political, demographic, and postal classifications (1996 boundaries).

Figure 4.2
Australian population density by Statistical Local Area, and State and Territory boundaries.

Figure 4.3
RRv disease notifications in Australia by year (July 1991 to June 1999).

Figure 4.4
Average annual RRv disease notification rates in Australia, by five year age group (July 1991 to June 1999).

Figure 4.5
National RRv disease notification rates per 100,000 by sex, for the years 1991/92 to 1998/99.

Figure 4.6
Statistical Local Areas shaded according to the Chi statistic. A value of zero indicates equilibrium; positive values indicate more RRv disease notifications than expected on the basis of population distribution; negative values indicate less than expected.

Figure 4.7
RRv disease notifications in Australia by month and year for States and Territories (July 1991 to June 1999).

Figure 4.8
RRv disease notifications in Australia, by month of onset (July 1991 to June 1999).

Figure 5.1
Major Australian rainfall zones.

Figure 5.2
Major Australian temperature and humidity zones.

Figure 5.3
Major Australian vegetation (Köppen) zones.

Figure 5.4
Major Australian rainfall zones, based on Statistical Local Areas.

Figure 5.5
Major Australian temperature and humidity zones, based on Statistical Local Areas.
Figure 5.6  Major Australian vegetation zones (Köppen), based on Statistical Local Areas.

Figure 5.7  Ross River virus bioclimatic regions, based on Statistical Local Areas (n=47).

Figure 6.1  Australia, showing the location of the Murray study area (shaded).

Figure 6.2  The two RRv bioclimatic regions of the Murray River (light shading), and the two Murray study regions contained within them (dark shading).

Figure 6.3  The Murray River study area: Regions 1 and 2. Major towns are highlighted.

Figure 6.4  Long-term average monthly rainfall and maximum and minimum temperature profiles for the Murray study regions.

Figure 6.5  Numbers of RRv disease notifications for each Murray study region, by year (July 1991 to June 1999).

Figure 6.6  Numbers of RRv disease notifications for Region 1 and Region 2 in the Murray, by month of onset (July 1991-June 1999).

Figure 6.7  1992/93 epidemic year: mean total rainfall for the two Murray study regions, by month.

Figure 6.8  1996/97 epidemic year: mean total rainfall for the two Murray study regions, by month.

Figure 6.9  Odds ratios for rainfall regressed against epidemics of RRv disease in the Murray regions.

Figure 6.10  Odds ratios for the number of rain days regressed against epidemics of RRv disease in the Murray regions.

Figure 6.11  Odds ratios for the Southern Oscillation Index regressed against epidemics of RRv disease in the Murray regions.

Figure 6.12  Odds ratios for Sea Surface Temperature regressed against epidemics of RRv disease in the Murray regions.

Figure 6.13  Odds ratios for average minimum temperature regressed against epidemics of RRv disease in the Murray regions.

Figure 6.14  Odds ratios for absolute minimum temperature regressed against epidemics of RRv disease in the Murray regions.

Figure 6.15  Odds ratios for average temperature regressed against epidemics of RRv disease in the Murray regions.

Figure 6.16  Odds ratios for average maximum temperature regressed against epidemics of RRv disease in the Murray regions.

Figure 6.17  Odds ratios for absolute maximum temperature regressed against epidemics of RRv disease in the Murray regions.

Figure 6.18  Odds ratios for relative humidity at maximum temperature regressed against epidemics of RRv disease in the Murray regions.

Figure 6.19  Odds ratios for relative humidity at minimum temperature regressed against epidemics of RRv disease in the Murray regions.
Figure 6.20 Odds ratios for vapour pressure regressed against epidemics of RRv disease in the Murray regions.

Figure 6.21 Amount of water released from the Hume Reservoir into the Murray River per week (1991 to 1999). Measured downstream at Doctor's Point, Albury, New South Wales.

Figure 6.22 Amount of water released from the Hume Reservoir into the Murray River per week (September to December 1996). Measured downstream at Doctor's Point, Albury, New South Wales.

Figure 6.23 Volume of water released from the Hume Reservoir and estimated land under inundation around Albury, New South Wales.

Figure 7.1 Australia, showing the location of the Southwest study area (shaded).

Figure 7.2 The Southwest RRv bioclimatic region (pale shading) and the Southwest study area within it (dark shading). Major towns in the study area are highlighted.

Figure 7.3 Statistical Local Areas within the Southwest study area.

Figure 7.4 Long-term average rainfall and maximum and minimum temperature profiles for the Southwest study region.

Figure 7.5 RRv disease notifications in the Southwest study region, by year (July 1991 to June 1999).

Figure 7.6 RRv disease notifications for the Southwest study region, by month of onset (July 1991 to June 1999).

Figure 7.7 Statistical Local Areas within the Southwest study area for which mosquito density data were available (coloured areas), and the approximate distribution of the mosquito trapping areas.

Figure 7.8 Odds ratios for the explanatory variables regressed against epidemics of RRv disease in the Southwest.

Figure 7.9 Odds ratios for the explanatory variables regressed against epidemics of RRv disease in the Southwest.

Figure 7.10 Odds ratios for the explanatory variables regressed against epidemics of RRv disease in the Southwest.

Figure 7.11 Monthly absolute tide heights, recorded at Bunbury, Western Australia (July 1991 to June 1999).

Figure 8.1 A simplified diagram for the use of indicators in an RRv disease early warning system.
Acronyms and abbreviations

ASGC  Australian Standard Geographical Classification
BFv   Barmah Forest virus
cm    centimetre
CSIRO Commonwealth Scientific and Industrial Research Organisation
ENSO  El Nino-Southern Oscillation
FAO   Food and Agricultural Organization
GIS   Geographic Information Systems
HIV/AIDS Human Immunodeficiency Virus/Acquired Immuno-Deficiency Syndrome
hPa   hectopascal
IgG   Immunoglobulin G
IgM   Immunoglobulin M
IPCC  International Panel on Climate Change
JE    Japanese encephalitis
km    kilometre
mm    millimetre
Murray Murray River study area
MVe   Murray Valley encephalitis
NCEPH National Centre for Epidemiology and Population Health
NDVI  Normalised difference vegetation index
NNDSS National Notifiable Diseases Surveillance System
RRv   Ross River virus
SLA(s) Statistical Local Area(s)
SOI   Southern Oscillation Index
Southwest South-west Western Australia study area
SST   Sea surface temperature
WHO   World Health Organization
Abstract

The purpose of this thesis is to determine if climate and environmental data could be used to predict epidemics of RRv disease with enough accuracy and lead time to be of use for public health surveillance. Diseases spread by mosquitoes cause extensive mortality and morbidity throughout the world. Climate is a primary influence on mosquito distribution, and on short-term and long-term disease trends. The current evidence suggests that, even if public health response is optimal, the combination and scale of climate change and other global changes will lead to an overall increase in mosquito-borne disease incidence in the future. The health impacts for communities will be strongly determined by the effectiveness of their public health system, and the adequacy of their response. It is therefore important to determine how public health authorities could improve the use of climate data for surveillance, analysis, and ultimately the prevention of mosquito-borne infections.

Ross River virus disease (RRv disease) is the main arbovirus of public health importance in Australia today. It is associated with a large burden of disease (on average 5000 cases of epidemic polyarthritis a year since 1991), and is likely to be an increasing problem in coming decades. Ross River virus (RRv) has a relatively complex transmission cycle, with multiple mosquito vectors and vertebrate hosts involved. Both long-term and inter-annual climate affect the distribution of the vectors, as well as the seasonality of transmission. Rainfall, temperature, humidity, and tidal inundation strongly regulate the transmission of the virus between the vector and vertebrate host populations.

The analysis of RRv disease notifications for Australia (1991 – 1999) showed that there was substantial variation in the seasonal and geographic disease trends across the country, which could not be explained by population effects alone. Analyses at the national and State levels, and even within broad scale rainfall zones, were too coarse to describe this variation, which was better revealed at the Statistical Local Area (SLA) level.
To enable an analysis of the climate and RRv disease relationship within biologically rational regions, a geographical classification system was purpose-built. SLAs were aggregated into “RRv bioclimatic regions” that were similar with respect to the major climatic and environmental factors that have been shown to influence RRv transmission.

From these regions, two areas of Australia were selected for intensive study. Predictive modelling was conducted in the Murray River (south-eastern Australia) and the Southwest (south-western Australia). Within each of these study areas, two models were developed. The “early warning” models used climate and environment variables for the months of July to November (austral winter to end spring) to predict the probability of epidemics. The “late warning” models included additional variables for the austral summer. In both study areas, a prerequisite for an epidemic related to the influence of mosquito breeding on vertebrate host immunity levels: rainfall in the preceding year’s spring to summer period was negatively associated with epidemics. In the Murray study area, the Region 1 early warning model had a sensitivity of 62% (specificity 95%), and the Region 2 model had a sensitivity of 73% (specificity 81%). For the late warning model, Region 1 had a sensitivity of 96% (specificity 93%), and Region 2 had a sensitivity of 66% (specificity 98%). In the Southwest, the early warning model had a sensitivity of 64% (specificity 96%). In addition to the predictive accuracy provided by the climate and environment data in this region, the value of mosquito surveillance data was also tested. These dramatically improved the sensitivity of the early warning model (sensitivity 90%, specificity 88%).

The results demonstrate that climate can be used as the basis for early warning of RRv disease epidemics with a lead time of several months – a substantial improvement in predictive certainty over what is currently available in most parts of Australia. Mosquito and vertebrate host surveillance is expensive, and only routinely conducted in very few parts of Australia. Climate data, on the other hand, are readily available and relatively cheap. This study used climate variables to account for the contribution of vertebrate host immunity to the occurrence of outbreaks.

The case studies illustrate the feasibility of establishing early warning systems for RRv disease. An early warning system comprises prediction and response. If effectively designed and implemented, it could enable improved timing of insecticide spraying, better communication of risk, and more sophisticated and specific prevention information. This would result in fewer cases of disease and a more efficient use of public health resources.
Chapter 1

Research context

1.1 INTRODUCTION

This thesis describes original research that was carried out on the relationship between climate and environment and the epidemiology of Ross River virus disease (RRv disease) in Australia. In Section 1.2 of this chapter, I argue that mosquito-borne diseases represent a major burden of disease at the global level and in Australia. Meteorological and environmental factors, amongst others, are major determinants of their distribution and occurrence. Climate forecasting can improve the public health management of mosquito-borne diseases by providing timely early warning of outbreaks. The rationale for focusing on RRv disease in this thesis is also presented. Section 1.3 is a description of the ecology of RRv, its vectors, and vertebrate hosts. Section 1.4 is an overview of the natural history of RRv in humans, and its distribution and determinants in populations. Section 1.5 assesses the evidence for the impact of weather and climate on the transmission of mosquito-borne diseases in general, and on outbreaks of RRv disease in particular. Studies that have successfully used meteorological and environmental factors to predict outbreaks of mosquito-borne diseases are reviewed. Section 1.6 is a summary of the chapter, and an outline of the gaps in research that form the basis for the work of this thesis.
1.2 THE BURDEN AND CONTROL OF MOSQUITO-BORNE DISEASES

Infectious diseases are “illnesses due to a specific infectious agent or its toxic products that have been transmitted from an infected person or animal reservoir to a susceptible host. This may occur directly, or indirectly through an intermediate plant or animal host, a vector, or the inanimate environment” (Last 2001).

Vector-borne diseases are a subset of this group, and encompass those infections that rely on a vector (such as mosquitoes, bats, flies, cockroaches, ticks, or rodents) to transmit the infection to humans. Sometimes animals or insects may also be involved as “reservoir hosts” in which the infectious agent normally lives and multiplies.

The major vector-borne diseases affecting humans at the global level – whether calculated by (i) the total population at risk of infection, (ii) the number of cases or people infected each year, or (iii) the number of life years lost to disease – are transmitted by mosquitoes (Table 1.1). Mosquitoes carry a number of disease-causing pathogens, including protozoa, worms and viruses. “Arboviruses” (arthropod-borne viruses) are a large group of viruses spread by certain invertebrates, most commonly blood sucking insects. Mosquitoes are one of the most common vectors for spreading arboviruses.

1.2.1 Main mosquito-borne diseases - global level

At the global level, malaria is considered the world’s most serious mosquito-borne disease (see Table 1). Approximately 40% of the world’s population is at risk of contracting the disease, and malaria is endemic in 92 countries (WHO 2002). In 1998 there were an estimated 1 million deaths globally from malaria, with 400 to 500 million cases of disease (WHO 1998a). Malaria is caused by four species of a protozoan parasite (plasmodium), transmitted between humans by the bite of infective female Anopheles mosquitoes. Of the four species that infect humans, two of them predominate overwhelmingly: Plasmodium vivax and P. falciparum. Vivax malaria, with its capacity for over-wintering dormancy, predominates in cooler, temperate zones. Falciparum requires warmer conditions and predominates in most tropical and sub-tropical regions. It has a much greater lethality than vivax malaria.
Table 1.1  Main vector-borne diseases, populations at risk and affected, and current assumed distribution.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Vector</th>
<th>Population at risk</th>
<th>People infected or new cases per year</th>
<th>Disability adjusted life years lost</th>
<th>Present distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malaria</td>
<td>mosquito</td>
<td>2400 million (40% world population)</td>
<td>273 million</td>
<td>39 million</td>
<td>Tropical and subtropical areas. High transmission areas on the fringes of forests in South America, south-east Asia (e.g. Thailand, Cambodia, Indonesia, Papua New Guinea), and parts of sub-Saharan Africa.</td>
</tr>
<tr>
<td>Schistosomiasis</td>
<td>snail</td>
<td>500-600 million</td>
<td>120 million</td>
<td>1.7 million</td>
<td>Africa; parts of South America; China, Taiwan, the Philippines, Indonesia, and the Middle East.</td>
</tr>
<tr>
<td>Lymphatic filariasis</td>
<td>mosquito</td>
<td>1000 million</td>
<td>120 million</td>
<td>4.7 million</td>
<td>Endemic in most of the warm, humid regions of the world.</td>
</tr>
<tr>
<td>Trypanosomiasis</td>
<td>fly</td>
<td>55 million</td>
<td>300-500 000</td>
<td>1.2 million</td>
<td>Tropical Africa.</td>
</tr>
<tr>
<td>Leishmaniasis</td>
<td>sandfly</td>
<td>350 million</td>
<td>1.5-2 million</td>
<td>1.7 million</td>
<td>Asia, Africa, Southern Europe, Americas.</td>
</tr>
<tr>
<td>Onchocerciasis</td>
<td>fly</td>
<td>120 million</td>
<td>18 million</td>
<td>1.1 million</td>
<td>Africa, Latin America, Yemen.</td>
</tr>
<tr>
<td>Chagas' disease</td>
<td>triatomine bug</td>
<td>100 million</td>
<td>16-18 million</td>
<td>600 000</td>
<td>Central and South America.</td>
</tr>
<tr>
<td>Dengue</td>
<td>mosquito</td>
<td>3000 million</td>
<td>tens of millions</td>
<td>1.8 million</td>
<td>Throughout all tropical countries. Most parts of Asia, the Pacific, Africa, the Caribbean and Central and South America, northern Australia.</td>
</tr>
<tr>
<td>Yellow fever</td>
<td>mosquito</td>
<td>508 million</td>
<td>200 000</td>
<td>not available</td>
<td>Africa, tropical South and Central America</td>
</tr>
<tr>
<td>Japanese encephalitis</td>
<td>mosquito</td>
<td>300 million</td>
<td>50 000</td>
<td>500 000</td>
<td>Western Pacific islands from Japan to the Philippines; many areas of eastern Asia from Korea to Indonesia, China and India; northern Australia.</td>
</tr>
</tbody>
</table>

Compiled from sources: (IPCC 2001a), (Chin 2000), (WHO 2003a).

Dengue is the most common arbovirus infection in humans. Humans are the main amplifying host of the virus, and dengue viruses are transmitted to humans through the bite of *Aedes* mosquitoes, the most important of which is the predominantly urban species *Ae. aegypti* (WHO 1993). There are four serotypes of the dengue virus. Dengue haemorrhagic fever and dengue shock syndrome are life-threatening complications that are thought to result from a second dengue infection, with a virus different in serotype to
that which caused the primary infection (Halstead et al. 1970). Prior to 1970, only nine countries had experienced dengue haemorrhagic fever epidemics (WHO 2003b). The global prevalence of dengue has grown dramatically in recent decades, and the disease is now endemic in more than 100 tropical and sub-tropical countries (South-east Asia and the Western Pacific are most seriously affected: WHO 2003b). There are in the order of 50 million cases of dengue infection worldwide every year, with an estimated 500 000 hospitalisations of dengue haemorrhagic fever (WHO 2003b). Mortality is difficult to estimate, as treatment can decrease the death rate from 20% (untreated) to 1% (with modern intensive care) (WHO 2003b). Some 138 000 deaths in the world were due to dengue haemorrhagic fever in 1996, caused by infection with dengue viruses of multiple serotypes (WHO 1997).

Yellow fever is a virus that is maintained in an enzootic cycle involving primates and mosquitoes. This can spill over into major epidemics when the urban vector *Ae. aegypti* is involved, an almost identical cycle to dengue (Monath 1988). Yellow fever is endemic in nine South American countries and in several Caribbean islands, and thirty-three countries are at risk in Africa. There are some 30,000 deaths from yellow fever per year (WHO 2003a). A vaccine is available that is safe and effective: a single dose provides protection for at least 10 years in 95% of people (WHO 2003a). Despite this, epidemics in tropical American cities have occurred in recent years because of low levels of population immunity (Gubler 1998a). If urban epidemics become more widespread in the tropical Americas as Gubler predicts (1998a), epidemics of the disease could possibly move to unvaccinated populations in cities in Asia and the Pacific.

A number of mosquito-borne viruses can cause encephalitis, with high case fatality rates (up to 60% for Japanese encephalitis, Murray Valley encephalitis and eastern equine encephalitis) depending on age, treatment availability, and the infecting agent (Chin 2000). Japanese encephalitis virus infects the largest number of people per year, and its distribution is throughout the western Pacific, from Japan to the Philippines, with sporadic outbreaks in eastern Asia (Chin 2000). Rift Valley fever, primarily a disease of domestic ruminants in parts of Africa, can also cause disease in humans, with complications that can be fatal. West Nile fever is common in parts of northern Africa, the northern Mediterranean area, and has recently become established in the United States (e.g. Eidson et al. 2001).
1.2.2 Main mosquito-borne diseases - Australia

A number of diseases spread by vectors other than mosquitoes affect people living in Australia (reviewed by Russell 1998a). These include diseases spread by ticks (tick paralysis and Queensland tick typhus), mites (scrub typhus), and fleas (murine typhus). However, none of these diseases are listed on the National Notifiable Disease Surveillance System, the cases are few, and there is minimal research, prevention, or control activities directed towards them. Consequently, mosquito-borne diseases are the most significant vector-borne diseases of public health concern in Australia.

Comprehensive reviews of the ecology of mosquito-borne diseases in Australia have been prepared by Mackenzie and others (1994b, 1998) and Russell (1995, 1998a). Those that contribute most to the current burden of disease are Ross River virus disease, Barmah Forest virus disease, Murray Valley encephalitis, dengue, and malaria. Japanese encephalitis, although not currently understood to be enzootic on the mainland, represents a potentially large threat to public health in the north.

1.2.2.1 Ross River virus disease

RRv disease is the most common arboviral disease in Australia. It also occurs in Papua New Guinea and the Solomon Islands (Flexman et al. 1998), and outbreaks have been reported in other Pacific island countries including American Samoa, Fiji, New Caledonia and the Cook Islands (Aaskov et al. 1981c, Rosen et al. 1981, Scrimgeour et al. 1987). Ross River virus causes epidemic polyarthritis (Fraser et al. 1986, Fraser and Marshall 1989, Flexman et al. 1998), which consists of arthritic symptoms that persist for several months and can be severe and debilitating (Harley et al. 2002, Mylonas et al. 2002). The disease is a significant public health issue in Australia, with 51 760 notifications from 1991 to 2002 (an average of some 4500 cases per year: Communicable Diseases Network Australia 2002). Recent estimates of the economic cost of the disease are in the order of $650 - $1265 per person (some A$3.1 and $6 million per year: Harley et al. 2001). There is no treatment for the disease, and mosquito control and public notification are the sole public health strategies.

The arbovirus is maintained by enzootic cycles involving vertebrate hosts and mosquitoes. Unlike some other mosquito-borne diseases, a wide range of mosquito species are believed to be capable of transmitting RRv, each species being adapted to different climate conditions (Russell 1995). Many different vertebrate hosts have been
suggested. The occurrence of RRv disease in Australia is widespread, with cases reported in all States. Notifications are reported throughout the year in the tropical northern parts of Australia (Kay and Aaskov 1989). Sporadic cases and occasional epidemics is the typical pattern in temperate southern Australia (Russell 1998a). The geographic distribution of notified RRv disease in the State of Queensland appears to have increased substantially since 1990 (Tong et al. 2001).

1.2.2.2 Murray Valley encephalitis

Murray Valley encephalitis (MVe) is a severe arboviral disease, transmitted by mosquitoes (predominantly Culex annulirostris). Historically, approximately one third of patients die, and about half the survivors suffer severe neurological effects, including paralysis and brain damage (Marshall 1989). MVe cases are occasionally reported in northern Western Australia, the Northern Territory, and northern Queensland, usually following heavy rain and flooding (for recent reviews, see: Russell 1998a, Spencer et al. 2001). In the south-eastern States, major outbreaks are believed to have occurred in 1917, 1918, 1951, and 1974, with almost no activity recorded in between (Forbes 1978). Murray Valley encephalitis is a relatively rare disease, although its increasing incidence (Spencer et al. 2001) and the severity of the condition makes it one of the higher mosquito-borne disease priorities in Australia.

The major vertebrate hosts are believed to be waterfowl, and the rufous night heron (Nycticorax caledonicus) has been shown to be particularly susceptible to experimental infection (Mackenzie et al. 1994b). Current understanding is that this virus has an area of enzootic activity in northern Western Australia (the Kimberley region), and is reintroduced into south-eastern regions by infected birds that migrate southwards following periods of extreme rainfall and widespread flooding (Spencer et al. 2001). In south-eastern Australia along the Murray River, outbreaks have been observed to occur mostly from January to April (austral summer and autumn), following widespread flooding over winter/spring (Forbes 1978). Forbes' hypothesis is also supported by the work of Nicholls (1986), who showed that the seven recorded epidemics of MVe were accompanied by low Southern Oscillation Index values in the preceding twelve months. He speculated that these La Niña seasons could provide objective early warning of the likelihood of MVe. However, as there has not been a major outbreak of MVe since 1974, and the climatic conditions predicted to precipitate such an event have also not occurred since then (Mackenzie et al. 1994c), the utility of these early warnings to public health
authorities (e.g, to enable timely increases in insecticide use and other prophylactic measures in years when Southern Oscillation Index values are low) is still not entirely clear. Even if Forbes' hypothesis is not completely reliable, it may be still useful as an aid, since there are few other predictive methods available (Bennett 2001).

1.2.2.3 Barmah Forest virus

Cases of Barmah Forest virus disease (BFv disease) were first recognised in New South Wales in 1986 (Vale et al. 1986), and the first outbreak of disease was recorded in Nhulunbuy in the Northern Territory in 1992 (Merianos et al. 1992). Since then information on the distribution and clinical symptomology of the disease has accumulated. In 1992, BFv disease became listed as a notifiable disease. From 1992 to 2002, some 7290 notifications of BFv disease were made, an average of 660 cases per year (Communicable Diseases Network Australia 2002). Clinical symptoms are similar to those of RRv disease (i.e., polyarthritis, fever, rash, headache: Flexman et al. 1998). The greatest number of notifications have been made for Queensland and New South Wales, with almost no activity recorded in South Australia or Tasmania (Communicable Diseases Network Australia 2002). BFv disease outbreaks have also been documented in Queensland (Phillips et al. 1990, Russell 1998a), and the south coast of New South Wales (van Buynder et al. 1995), as well as south-west Western Australia (Lindsay et al. 1995) and Victoria (Passmore et al. 2002). Russell (1998a) has suggested that BFv activity appears to be spreading, and cites the relatively recent first finding of BFv in Western Australia in 1992 (Broom et al. 1994), and the lack of evidence from animal serosurveys and mosquito collections in that region prior to that time to support his view.

1.2.2.4 Malaria

Approximately 8726 (imported) malaria cases were notified in Australia between 1991 and 2002 - an average of 790 per year (Communicable Diseases Network Australia 2002). All States report imported cases, slightly more than one half of which occurred in the Northern Territory and Queensland. The rate of notifications has steadily increased since the 1960s, reflecting the increase in travel between Australia and endemic regions.

Malaria eradication in Australia was achieved in the 1960s (Bryan et al. 1996). Only twelve locally acquired cases of malaria have been reported in Australia since 1962, all in Far North Queensland (Walker 1996, Brookes 1997). Most notable was an outbreak of vivax malaria at Cape Tribulation in October 2002 (Scott Ritchie pers. comm. 2002), in
which ten people were locally infected. *Anopheles farauti sensu lato* are the only mosquitoes conclusively shown to have transmitted malaria in Australia (in an outbreak in Cairns in 1942: Walker 1998a), and observations since the 1900s suggest they are the most important Australian vectors (Sweeney *et al.* 1990, Cooper *et al.* 1995, Bryan *et al.* 1996).

### 1.2.2.5 Dengue

Dengue fever is transmitted in mainland Australia by the freshwater breeding mosquito *Ae. aegypti* (Sutherst 1994). Imported cases are regularly diagnosed in all capital cities. 2776 cases of dengue were recorded in the period 1991-2002, approximately 250 cases per year (Communicable Diseases Network Australia 2002). Local transmission following the importation of virus occurs occasionally, and since the 1940s such cases have been confined principally to the northern and eastern parts of Queensland (Mackenzie *et al.* 1996a). In these cases, infection typically spreads from a traveller via mosquitoes to local residents. The large numbers of tourists that now travel between Australia and countries in Asia and the Pacific, where dengue is endemic, has greatly increased the risk of introduction of the virus (McBride 1999). Several large outbreaks have occurred in recent years: in Charters Towers in 1993 (26% of the non-immune population were infected: McBride *et al.* 1998), in Cairns and the Torres Strait in 1996 (201 cases), and in the Cairns region in 1998 (500 cases: Hanna *et al.* 2001).

The principal dengue vector, *Ae. aegypti*, is currently established in the north and central areas of Queensland (Mackenzie *et al.* 1996a), although the virus is not believed to be endemic in Australia at present. *Ae. aegypti* was once more widely distributed throughout Australia in the past, and outbreaks of dengue were reported to be common in Brisbane, and even extended into Sydney before WWII (Lee *et al.* 1989). Key factors in the emergence and then later disappearance of the vector from such sub-tropical and temperate regions have been the increasing urbanisation of society, in particular the conversion from rainwater tanks to a reticulated water supply, the use of refrigerators instead of water-cooled safes, diesel- rather than steam-powered trains, and the use of domestic insecticides (Mackenzie *et al.* 1996b, Reiter 1996, Gubler 1998b).

In a risk assessment of the health impacts of future climate change in Australia, McMichael and others (2003) concluded that a number of factors combine to make dengue a greater public health threat in Australia than malaria. First, there is more
potential for dengue outbreaks to spread rapidly within populations, as treatments are not available to reduce the period of viraemia. Effective, fast-acting treatments are available for malaria, and thus malarious people remain infectious for a much shorter period. Second, the breeding and feeding activities of the dengue vector (a morning/evening biter that prefers to breed in the urban environment and to feed on humans) require more complex and systematic prevention efforts than those needed to control malaria vectors (which do not breed so close to urban environments and generally bite at night - thus bed nets provide a simple form of protection in affluent countries). For these reasons, the establishment of the vector in Australia and an individual's risk of infection are likely to be greater with dengue.

1.2.2.6 Japanese encephalitis

The first outbreak of Japanese encephalitis (JE) occurred in 1995 on Badhu Island in the Torres Strait region in northern Australia (3 cases were reported, two of whom died: Hanna et al. 1996). Since then, a second outbreak was reported on Badhu Island in 1998, and in April of that year the first case on the Australian mainland was diagnosed (CDI 1998). Sentinel animal sites (pigs) have been established to monitor JE activity in the far north region (under the direction of the Northern Australian Quarantine Strategy). The vertebrate hosts for JE (wild pigs and waterbirds) exist in sufficient number and in proximity to the main Australian vector, Cx annulirostris (Mackenzie 1999). This fact, coupled with the wetland habitat of the northern Australian region, increases the potential for the virus to establish in an enzootic cycle, and the further possibility of it moving south to more populous areas in the future (Mackenzie 1999).

1.2.3 Determinants of mosquito-borne disease transmission

A complex interplay of social, economic, political, and biophysical factors (Daily and Ehrlich 1996) shape the patterns of mosquito-borne diseases. Global changes that impact upon the biophysical environment and increase human vulnerability include alterations to land use, increasing consumption and resource usage, climate change, and the growth and re-distribution of human populations (McMichael 1993, IPCC 2001a). Winch (1998) has argued that most vector-borne diseases are diseases of poverty, and the association between malaria and other infectious diseases with human rights violations (particularly land removal, deforestation, enforced migration) has often been observed (Pederson 1996). Population and individual-level ability to protect against mosquito-borne diseases is also challenged by changes in water management (such as dam construction, irrigation,
and other water development projects: Gratz 1999), conflict and wars, the loss of biodiversity and indigenous knowledge, immune suppression, the increase in insecticide and antibiotic resistance, poor public health infrastructure, and a shift in emphasis from prevention to emergency response (Daily and Ehrlich 1996, Gubler 1996, Molyneux 1997, Gubler 1998b, Githeko et al. 2000). For example, proximal geographic regions that share the same climate and weather, but have widely disparate economic resources, can exhibit significantly different rates of dengue infection (Reiter et al. 2003).

Notwithstanding these influences, the heterogeneity of climates and landscapes broadly determines the distribution of mosquito-borne diseases (Bergquist 2001), and weather is a major determinant of incidence. Weather factors such as temperature, rainfall, and humidity are capable of assisting or interrupting the biology and population dynamics of vector mosquitoes (Reeves et al. 1994), thereby influencing distribution, incidence and severity of disease transmission. Rainfall (or lack of it) plays a crucial role in the epidemiology of arboviral diseases as it provides the medium for the aquatic stages of the mosquito life cycle. Temperature impacts on mosquito productivity and on viral replication. Humidity affects mosquito survival, and hence the probability of transmission (Sellers 1980, Reiter 1988, Leake 1998).

For example, the relationship between climate and malaria distribution, and between weather and variations in malaria transmission, has been well documented (for example see Martens et al. 1995, Lindsay and Birley 1996, Martens 1998, IPCC 2001a, Kovats et al. 2001). Seasonal peaks in dengue transmission, as well as the geographic boundaries of transmission, have been related to high rainfall and humidity (IPCC 2001a, Hales et al. 2002a).

1.2.4 Climate as a tool for early warning of mosquito-borne diseases

Mosquito control and public notification remain the only public health response to the majority of mosquito-borne diseases (with the exception of yellow fever and Japanese encephalitis, where vaccination is possible although not widely used: Gubler 1998a). Both measures require knowledge of an impending outbreak, and a suitable response time. Climate factors that drive the growth of mosquito populations and the replication of pathogens have the potential to be used as a proxy for early warning of the probability of an outbreak of disease. The objective is to detect conditions suitable for pathogen
amplification in the natural cycle at the earliest possible time (Rose et al. 1999), so that public health interventions can have the greatest opportunity for success.

Climate forecasts will be helpful tools in the public health management of mosquito-borne diseases if they can (i) improve the targeting and sensitivity of surveillance and increase the length of the response time above existing surveillance activities (Linthicum et al. 1999), or (ii) reduce the cost of traditional surveillance activities.

1.2.5 Rationale for focusing on Ross River virus disease

Mosquito-borne diseases represent a major public health problem in Australia and elsewhere. Although there are multiple determinants of individual outbreaks, climate and weather together contribute substantially to the distribution and level of these diseases. Climate and weather are also highly amenable to predictive modelling, with the purpose of improving the timeliness of early warning of disease events.

Although a number of mosquito-borne diseases are of public health importance in Australia, I chose to restrict the studies of this thesis to RRv disease for several reasons.

- The incidence of the disease is sufficiently high for the influence of climate to be revealed. In the period July 1991 to June 1999, 39,422 notifications of RRv disease were recorded for the country as a whole (Communicable Diseases Network Australia 1999). This number provided sufficient statistical power to undertake investigations at both national and regional levels.

- The large number of cases of RRv disease - on average, more than 5000 per year during this period - indicates that the disease has a significant economic and social impact on the nation, although this is still to be comprehensively estimated. The numbers of people infected annually, the morbidity associated with the disease, and the wide geographic spread of cases (reported in every Australian State and Territory), suggest that RRv disease is a considerable public health issue.

- The ecology of the virus, and of its vector and vertebrate host species, has been comparatively well documented.

- A scientific advantage for studying RRv disease is that it provides an excellent ecological experiment. The main vectors of RRv breed in salt marsh or inland freshwater pools and streams, rather than domestic environments. Many local governments in Australia do not conduct routine insecticiding or mosquito habitat destruction. As a consequence, there is remarkably little human "interference" with
breeding cycles, compared to other mosquito-borne diseases such as dengue or malaria.

- Human contact with the RRv natural cycle is increasing in many parts of Australia. Some of this is due to increases in mosquito populations as a result of irrigated agriculture in otherwise arid inland areas (as has been the case in the Griffith region, New South Wales), to tourism (as in parts of the Loddon Mallee Region, Victoria), or to residential developments in the vicinity of mosquito breeding habitats (as in parts of Brisbane City Council, and in the Peel Inlet region of south-west Western Australia). Future changes to climate are, in general, predicted to expand the transmission of RRv to new parts of Australia (Lindsay and Mackenzie 1997), and to lengthen the mosquito breeding season in some regions (Russell 1998b).

1.3 Ecology of Ross River Virus

Reviews of the ecology of RRv have been prepared by a number of authors (for example Marshall and Miles 1984, Kay and Aaskov 1989, Mackenzie et al. 1994b, Russell 1994, Mackenzie et al. 1998, Russell 1998a), and it is beyond the scope of this thesis to cover that work in detail. The purpose of this section is to describe those aspects of the ecology of RRv relevant to the prediction of outbreaks of human disease. In this context, an understanding is needed of the types of vectors and vertebrates involved in the transmission of the virus, and the mechanisms for virus persistence in the environment. Details of the impact of weather and climate on vector, vertebrate, and virus life cycles are discussed in a later section.

1.3.1 Ross River virus

RRv is an arbovirus, one cause of epidemic polyarthritis in people. It can be further classified as an alphavirus from the family Togaviridae.

1.3.1.1 The RRv transmission cycle

The transmission of arboviruses requires an adequate supply of competent vectors, virus, and susceptible vertebrate hosts in a locality at the same time. Vertebrate hosts are those in which the virus normally lives, and on which it depends primarily for survival and reproduction (Last 2001). Synonymous terms include "reservoir host" and "amplifying
host. Arbovirus transmission can be either vertical (through infection of eggs to the next generation of vector) or horizontal (following replication of virus in the vector, the female arthropod bites a susceptible vertebrate host and transmits the virus) (Turrell 1989).

The transmission cycle for Ross River virus is relatively complex, in that multiple vertebrate hosts and vectors are involved. Humans do not form part of the natural cycle, but appear able to act as both dead end hosts and amplifying hosts in epidemic situations. The virus multiplies to produce viraemia in the vertebrate host, and the mosquito vector becomes infected by feeding on the host. Following ingestion, the virus undergoes a developmental phase in the mosquito where it moves from the mid-gut through other organs and into the salivary glands and saliva. After this period of extrinsic incubation, the virus can be passed on to new vertebrates by the bite of the mosquito (Leake 1998). Once a host is infected, the viraemic phase for RRv usually lasts several days to a week (Kay and Aaskov 1989), after which the host develops immunity and is removed from the population of infectives. Infection of humans can occur when the abundance of infected vectors expands explosively into major epizootic disease activity, and humans become involved as alternative hosts in the transmission cycle. “Spill-over” occurs when the virus is transmitted from reservoir populations to humans (Daszak et al. 2000).

The dynamics of RRv transmission vary in different parts of the country. Transmission of all arboviruses is highly seasonal, dictated principally by environmental temperature, the availability of water-filled breeding sites for the aquatic life stages of the mosquito, and the population dynamics and host-biting preferences of the mosquitoes (Leake 1998). In the tropical and sub-tropical regions of Australia, temperature and precipitation levels enable adult RRv vector species to remain active all year (Kay and Aaskov 1989) and cases are reported in most months of the year. Even if mosquitoes bite humans only infrequently in the presence of other (preferred) animals, their large numbers year-round in these regions enable them to maintain viral transmission (Kay et al. 1987). In colder, temperate regions, mosquitoes are active only during the warmer months (spring to summer) (Dhileepan 1996), and viral activity is intermittent.

1.3.1.2 Endemic and epidemic cycles of transmission

In this thesis the terms “endemic” and “epidemic” refer to the human epidemiological patterns of disease occurrence, rather than to the ecological pattern of the persistence of
the RRv. Where cases of disease are reported constantly throughout the year over a long period at about the same incidence, this is referred to as an endemic (Last 2001). In the case of RRv disease, endemicity can occur in regions where the virus is enzootic (i.e., where there is continued and constant presence of the virus in the animal population over a long period of time: McMichael et al. 1996). Endemic regions for RRv develop in areas where humans intrude permanently into the natural cycle (such as the Peel Inlet area of Western Australia, or outer suburbs of Brisbane), or where the enzootic zone is expanded to include the human habitat (as in the irrigated areas of the Ord River in Western Australia).

The term "epizootic" refers to an outbreak of disease in an animal population (Last 2001). Epizootic regions are those where circulation of the virus in the animal population is limited to only some months of the year, or to infrequent years. An epidemic occurs when cases of disease are in excess of normal expectancy (Last 2001). An epidemic can occur in regions where human cases of disease are rarely recorded, or in endemic regions, when large increases from the background (endemic) level occur. As I explain later, the term "epidemic" is used in this thesis to specifically refer to the outcome variable for modelling. To distinguish from that particular use of the term, other discussions of large occurrences of RRv disease in a confined place and time are hereafter referred to as "outbreaks".

1.3.1.3 Virus persistence and re-introduction mechanisms

Without the means of viral reintroduction, the transmission of arboviruses within epizootic regions would not occur (Danielova 1975). In temperate climatic zones, where the circulation of arboviruses is limited to a few months during the summer (warmer) activity of mosquitoes, and in arid regions where precipitation conditions are unfavourable for continued mosquito breeding (Danielova 1975), the mechanism for arbovirus persistence in the environment is not well understood. Various mechanisms have been flagged for the survival of Australian arboviruses, listed at Table 1.2. Of these, Lindsay (1993b) has suggested that the survival of female mosquitoes over the cooler months (overwintering) and vertical transmission (survival of virus in the eggs of the mosquito) are the two primary mechanisms for Ross River virus.
Table 1.2 Proposed mechanisms of survival for Australian arboviruses.

<table>
<thead>
<tr>
<th>Survival mechanism</th>
<th>Proposed by</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Year round transmission in wet tropical areas</td>
<td>(Doherty et al. 1979)</td>
</tr>
<tr>
<td>• Over-wintering in old adult female mosquitoes</td>
<td>(Russell 1986b)</td>
</tr>
<tr>
<td>• Reintroduction to the region from migrating vertebrates</td>
<td>(Forbes 1978, Coelen and Mackenzie 1988, Broom et al. 1995)</td>
</tr>
<tr>
<td>• Movement of virus from one geographic region to another, via viraemic humans or livestock</td>
<td>(Marshall and Miles 1984, Lindsay et al. 1992)</td>
</tr>
<tr>
<td>• Vertical transmission of virus from one generation to the next, via drought resistant eggs (and possibly larvae)</td>
<td>(Kay 1982, Broom et al. 1995, Dhileepan et al. 1996)</td>
</tr>
</tbody>
</table>

Adapted from: (Lindsay et al. 1993b)

Vertical transmission (or transovarial transmission) through generations of mosquitoes has been demonstrated within the laboratory and is gaining favour as a hypothesis of persistence for RRv. The multiplicity of RRv vectors and vertebrate hosts; the short extrinsic incubation period in vectors; and desiccation-resistant vector eggs (Kay et al. 1987, Kay and Aaskov 1989, Lindsay et al. 1993b) are all factors that support this hypothesis.

1.3.1.4 Geographic distribution of the virus

The combined evidence of a number of studies suggests there is geographic strain variation in RRv, which supports a view that discrete vector and virus cycles operate in different parts of the country. For example, the work of Wells and others (1994) suggests that the virus can adapt to local vertebrate host populations. Virulence studies in mice have shown variations in virus isolates across the country (Kay and Aaskov 1989). Lindsay and others (1993c) used RNase T1 maps to compare different isolates of RRv. They suggested that two major genetic topotypes of the virus exist: one in the east and one in the west, which meet and overlap in the Kimberley region of north-west Australia. In later work, Sammels and others (1995) conducted nucleotide sequences of the virus and further categorised the eastern genetic type into north-eastern and south-eastern. Isolates of RRv from the Pacific Island were identified as belonging to the south-eastern genotype (Sammels et al. 1995).
1.3.2 Vector species

Ross River virus has been isolated from 28 species of wild caught mosquitoes from the genera *Aedes*, *Culex*, *Anopheles*, *Coquilletidia*, *Mansonia*, and *Tripteroides* (Russell 1998a). What this means with respect to the capacity of these species to effectively transmit the virus is largely unknown, except for a few major species. *Cx. annulirostris* is generally considered the most important inland vector, and *Ae. vigilax*, and *Ae. camptorhynchus* the most important coastal vectors (Campbell et al. 1989, Mackenzie et al. 1994b, Russell 1994, Russell 1995, Russell 1998a). Which species is responsible for the transmission of RRv depends on the geographical region, the month of the year, and the climate and environmental conditions (Marshall and Miles 1984).

Further, the composition of vectors in a region and the efficiency of vectors involved in different stages of the transmission cycle influence the level of RRv activity (Dhileepan 1996), and hence the likelihood of human infections. The host preference of the vector species also indicates whether it is responsible for amplifying the virus in the natural cycle, or for transmitting the virus to humans. Some vector species are only zoophilic in habit (i.e., they feed exclusively on animals other than humans), whereas others are primarily zoophilic but “during periods of peak transmission a small proportion of these host-seeking mosquitoes may opportunistically bite humans” (Leake 1998).

The following sub-sections provide an overview of the distribution, host preferences, and breeding patterns of the major RRv vector species (Table 1.3). Greater detail of species is provided in Chapters 6 and 7.

---

1 During the latter course of this PhD, a taxonomic revision of the genus *Aedes* and its subgenera occurred, with the result that several subgenera have been raised to generic rank and have changed name from *Aedes* to *Ochlerotatus* (NSW Arbovirus Surveillance & Vector Monitoring Program 2003a). For ease of citing references, and to avoid confusion with previous published work, I have retained the use of the older classification system throughout this thesis (i.e., *Aedes*).
Table 1.3 Major mosquito species implicated as vectors of Ross River virus.

<table>
<thead>
<tr>
<th>Mosquito species</th>
<th>Host preferences</th>
<th>Distribution</th>
<th>Breeding habitat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cx annulirostris</td>
<td>Humans, other mammals (occasionally birds)</td>
<td>Throughout mainland Australia</td>
<td>Shallow, freshwater pools</td>
</tr>
<tr>
<td>Ae. vigilax</td>
<td>Mammalian vertebrates (occasionally birds)</td>
<td>Generally northern coastal areas (tropical and sub-tropical); inland saline areas</td>
<td>Salt water; a floodwater species that lays eggs on the edge of salt marshes and mangrove wetlands</td>
</tr>
<tr>
<td>Ae. camptorynchus</td>
<td>Mammalian vertebrates (occasionally birds)</td>
<td>Generally southern coastal areas; inland saline areas</td>
<td>Salt water; brackish paperbark swamps</td>
</tr>
<tr>
<td>Ae. notoscriptus</td>
<td>Humans, among a variety of mammals</td>
<td>Northern Australia</td>
<td>Peri-domestic container breeder</td>
</tr>
</tbody>
</table>

References:
1. (Kay 1979b) 7. (Marshall and Miles 1984) 13. (Lee et al. 1984)
5. (Harley et al. 2000) 11. (Standfast and Barrow 1969) 17. (Russell 1998a)

1.3.2.1 Culex annulirostris

The freshwater breeding Cx annulirostris is commonly distributed throughout tropical and sub-tropical Australia (Dale et al. 1998). On the basis of virus isolation data, Cx annulirostris has been implicated as the major transmission vector of RRv in the Northern Territory (Whelan et al. 1996), in northern Queensland (Harley et al. 2000) and in parts of northern Western Australia (Mackenzie et al. 1994a). It is also common throughout southern Australia, and is considered to be the major RRv transmission vector in the south-eastern temperate and semi-arid inland areas of New South Wales, Victoria, and South Australia (Marshall and Miles 1984, Kay et al. 1986, Kay and Standfast 1987, Russell 1994). In the Murray Valley region of Victoria, for example, Cx annulirostris density was correlated with a widespread outbreak of RRv disease in 1992/93 (Dhileepan 1996).

Shallow, transient freshwater pools are the preferred site of egg-laying for this species (Reeves et al. 1954, Standfast and Barrow 1969), as persisting pools attract predator populations and support less suitable vegetation (McDonald 1980). As a result, the greatest population increases of Cx annulirostris are often observed after heavy rainfall, when temporary rain-filled pools and flooded grass lands are created, and in irrigated rice fields and dams (Dobrotworsky 1965, Kay and Aaskov 1989). It is a prolific and
opportunistically, and can exploit shallow grassy pools within one day of their formation (McDonald and Buchanan 1981).

Breeding occurs from mid-spring to late autumn in the southern temperate regions (Dobrotworsky 1965), and populations typically peak in January or February (Russell 1986b). The size of the adult population reflects the availability of larval habitat in summer, and consequently adult populations are very low in drought years (Russell 1986b). In northern Australia, peak abundance has been recorded in Queensland during the mid and late wet season (Kay 1979a), and around Darwin in the Northern Territory following the end of the wet season (Russell 1986a).

_Cx. annulirostris_ feeding preferences include humans, other mammals, and occasionally birds (Kay 1979b, Lee _et al._ 1989). Clonan and Russell (1986) noted a possible seasonal "shift" in host attraction from birds in early summer to mammals in later summer for this species. Mark-release-recapture techniques have demonstrated the species is capable of dispersing an average distance of 4.4 km in one study (Bryan _et al._ 1992) and 6.8 km in another (O'Donnell _et al._ 1992). In New South Wales and Victoria, it is thought to overwinter as quiescent adults (Russell 1986a, Clancy and Russell 1994), enabling virus survival in these temperate regions.

### 1.3.2.2 _Aedes vigilax_ and _Aedes camptorynchus_

The salt-water species _Ae. vigilax_ and _Ae. camptorynchus_ have both been shown to be competent laboratory vectors of RRv (Kay 1982, Ballard and Marshall 1986). _Ae. vigilax_ is considered the dominant salt-water vector for RRv in the more northern (i.e., tropical and sub-tropical) parts of coastal eastern and western Australia (Kay _et al._ 1986, Dale _et al._ 1998), and has been noted as an important vector in the Northern Territory in the early wet season (Tai _et al._ 1993, Russell 1995). Isolates of RRv have been confirmed from _Ae. vigilax_ as far south as New South Wales (Russell and Clonan 1989, Russell _et al._ 1991). In this region, virus activity was noted in _Ae. vigilax_ from midsummer through to autumn during 1985-1988, which corresponded with confirmed cases of RRv disease in the region (MJ Clonan unpublished data, reported in Russell _et al._ 1991).

_Ae. camptorynchus_ on southern Australian coastlines occupies a similar ecological niche as _Ae. vigilax_ on more northerly coastlines (McManus and Marshall 1986). _Ae. camptorynchus_ is considered the major transmission vector of RRv in south-west Western
Australia (Kay and Aaskov 1989, Lindsay et al. 1989, Mackenzie et al. 1994c), in coastal Victoria (Dhileepan et al. 1996), and in Tasmania (McManus and Marshall 1986).

*Ae. camptorynchus* and *Ae. vigilax* are floodwater species that lay their eggs on the outer edges of salt marshes, which are usually dry. After the flooding of egg sites by spring tides (in the temperate south of the country) or by heavy summer rainfall over the winter and spring months (in the semi-arid and temperate inland areas of the south-east), mosquito larvae hatch in 7-10 days and massive population growth can follow (Kay and Aaskov 1989). *Aedes* species usually live in coastal marsh and mangrove wetlands in the tropical north (Russell 1998a) or in the brackish paperbark swamps of the cool temperate south (McManus and Marshall 1986). However, they can also become permanently established inland wherever brackish water exists (Dobrotworsky 1965). Large populations of both *Ae. camptorynchus* and *Ae. vigilax* inhabit the saline areas of the Murray River region of south-eastern Australia, occupying inland temperate and semi-arid niches (Dobrotworsky 1965, Kay and Aaskov 1989, Lee et al. 1989). The eggs of these species are desiccation-resistant, and the virus can survive in this way for several years, even in drought conditions (Kay and Aaskov 1989). In the coastal temperate regions of Victoria, the mechanism for virus persistence is believed to be due to the capacity of *Aedes* species to overwinter as larvae, with fluctuations in populations being triggered by tidal inundation (Dhileepan et al. 1996).

*Ae. camptorynchus* and *Ae. vigilax* have been observed to prefer mammalian vertebrates to birds, although they will feed on both (Lee et al. 1984, Cloonan and Russell 1986). *Ae. vigilax* is most commonly an evening biter, and rarely bites indoors in Australia (Lee et al. 1984). *Ae. camptorynchus* will bite during the day, but peak biting also occurs after sunset (Lee et al. 1984).

### Other possible RRv vectors

There is increasing evidence that *Ae. notoscriptus*, a peri-domestic container breeder, may be an important urban vector. RRv was isolated from field caught *Ae. notoscriptus* in Darwin in 1983 (Whelan and Weir 1983) and in Brisbane in 1994 (Ritchie et al. 1997). The virus has also been isolated from *Ae. notoscriptus* in laboratory studies (Watson and Kay 1998). In the latter, transmission occurred between 9 and 14 days after infection, with a maximum rate of 13% between days 12 and 14, which suggests that although it is a possible vector of RRv it may only have low efficiency (Watson and Kay 1998). However,
the fact that it is known to feed on humans, among a variety of mammals (Ritchie et al. 1997), its peri-domestic breeding habits, and its day and night time feeding patterns (Lee et al. 1982), means that its role in urban transmission cannot yet be discounted.

High densities of *Cx australicus* have been observed spatially coincident with large outbreaks of RRv disease in the Murray Valley region in 1992/93 (Dhileepan 1996). RRv was also isolated from wild-caught male *Cx australicus* in arid Western Australia in 1992 (Lindsay et al. 1993b). As this species does not normally feed on humans (McDonald 1980, Russell 1993) it is unlikely to be involved in the infection of humans with RRv. However, due to its abundance during spring and autumn in the Murray Valley (Dhileepan 1996) it has been implicated in the early season amplification of RRv among vertebrate hosts (Russell 1993).

Gubler (1981) conducted laboratory studies to determine the transmission efficiency of *Ae. polynesiensis* and *Ae. aegypti* for RRv. Both species were found to be efficient vectors, with *Ae. polynesiensis* being the most susceptible. RRv has not been isolated from either of these species in the field (Gubler 1981, Kay and Aaskov 1989). However, if *Ae. polynesiensis* was found to be a competent vector and the virus was introduced into the region, this species' distribution throughout the south and central Pacific islands, and its close association with human habitation, could make RRv disease a widespread public health problem (Gubler 1981).

A range of other vectors have been suggested as playing a role in small epidemic outbreaks, or in maintaining the natural cycle in various parts of the country (Russell 1994). Other than the species mentioned, those that have been implicated on the basis of virus isolations or experimental infection from female mosquitoes include *Anopheles annulipes* (Marshall and Miles 1984, Lindsay et al. 1989), *Coquillettidia linealis* (Cloonan and Russell 1986, Kay and Standfast 1987, Russell and Cloonan 1989, Jeffrey et al. 2002), *Ae. alternans* (Cloonan and Russell 1986, Russell and Cloonan 1989, Ritchie et al. 1997), *Ae. normanensis* (Broom et al. 1989), *Ae. procax* (Ritchie et al. 1997), *Ae. funereus* (Ritchie et al. 1997, Ryan et al. 1997), *Ae. flavifrons* (McManus and Marshall 1986), *Cx sitiens* (Lindsay et al. 1993b, Ritchie et al. 1997), and *Cx quinquefasciatus* (Lindsay et al. 1993b). Although RRv has been isolated from this wide range of mosquito species, their vector competence and role in the RRv transmission cycle is still unknown (Mackenzie et al. 1994c), and as such their vector status is still to be determined.
1.3.3 Reservoir hosts

1.3.3.1 Evidence implicating vertebrate hosts

A substantial body of experimental studies has been conducted to ascertain vertebrate host competency for RRv (including work by Doherty et al. 1966, Doherty et al. 1971, Gard et al. 1973, Marshall et al. 1980, Marshall and Miles 1984, Kay et al. 1986, McManus and Marshall 1986, Kay and Standfast 1987, Campbell et al. 1989, Vale et al. 1991, van Buynder et al. 1995). In a review of this work, Kay and Aaskov (1989) noted that nineteen species of birds and mammals have been shown to develop at least a transitory viraemia to RRv, including kangaroos, wallabies, sheep, pigs, horses, rabbits, mice and birds. They concluded that marsupials are more competent vertebrate amplifiers of Ross River virus than placental mammals, which are in turn more effective than birds (Kay and Aaskov 1989).

Recently, Harley and others (2001) have challenged this conclusion, suggesting that knowledge of RRv vertebrate hosts is rather uncertain, and thus it is premature to rank animals for their importance as reservoirs of the virus. They noted that although Kay and Aaskov's review (1989) was based on a large number of antibody prevalence studies, only seven RRv isolate studies were involved, and no consideration had been given by the authors to other vector competence criteria (such as information on behaviour, ecology, or reproductive cycles). Harley and others (2001) conducted an updated review of the evidence. To summarise, their findings were that (i) sufficient data exist to support the hypothesis that macropods (wallabies and kangaroos) are the major vertebrate hosts of RRv in non-urban environments, (ii) as macropods are rare in urban settings (except at the fringes), other animals are likely to be the primary reservoirs in metropolitan areas, and (iii) there are good grounds for reappraising the role of birds as RRv reservoir hosts given their ubiquity, the low number of species studied, and the possibility that they only have a short period of viraemia. It is worth noting that Mackenzie (1999), Mackenzie and Smith (1996), and Russell (1994, 1995) have also considered that macropods are the major vertebrate hosts of the virus in non-urban settings.

Work by Boyd and others (2001) has confirmed the long-suspected importance of the Australian brushtail possum (Trichosurus vulpecula) as a reservoir host in the urban transmission cycle. Horses have also been implicated as potential amplifying reservoirs in metropolitan outbreaks when they are present in large numbers, on the basis of virus
isolations (Pascoe et al. 1978, Campbell et al. 1989) and high-titred viraemias (Kay et al. 1986). They are also the only animals that have been demonstrated to be affected clinically (Kay and Aaskov 1989). Domestic dogs and cats, although exposed naturally to RRv and capable of developing antibodies, are unlikely to be important reservoirs of the virus (Boyd and Kay 2002).

Humans may also be capable of effectively amplifying RRv in some situations: the outbreaks of RRv disease in the Western Pacific region during 1979/1980 demonstrated that humans have a period of viraemia prior to the development of antibodies, and can be involved as vertebrate hosts in epidemic events (Aaskov et al. 1981c). Rosen and others (1981) isolated Ross River virus with high titres from patient sera up to seven days after onset in the Cook Islands. Outbreaks in the city of Perth, Western Australia, in 1991/92, were also believed to have been maintained by humans, in addition to other species common to those environments (Lindsay et al. 1992). To date, however, the extent to which humans are able to effectively amplify the virus (i.e., increase the amount of virus in the environment) is still unknown.

1.3.3.2 Distribution and breeding cycle of the main reservoir host species

For reasons discussed more fully in Chapter 5, the analyses of climatic and environmental influences on RRv disease epidemics were constrained to the rural regions of Australia. In light of this, and based on the evidence presented above, the following section discusses the distribution and breeding cycle of the major rural reservoir hosts for RRv - macropods (kangaroos and wallabies). To date there has been almost no research undertaken on the ecology of RRv in these vertebrate hosts. Of relevance to the studies of this thesis are questions about the length of the kangaroo breeding cycle, its relation to host immunity levels, and to the length of the RRv transmission cycle in epizootic regions.

The three large kangaroos - the Red kangaroo (Macropus rufus), Eastern Grey kangaroo (M. giganteus) and Western Grey kangaroo (M. fuliginosus) - belong to the family Macropodidae, which includes 54 species that range in size from large kangaroos to small rat-kangaroos (Strahan 1991). These three species live in Australia's pastoral rangelands and arid interior, and are the most plentiful of all the macropod species (Figure 1.1 illustrates their approximate range). Red kangaroos are considered abundant over much
Figure 1.1 A simplified representation of the distribution of the three major kangaroo populations in Australia.

Adapted from: Southern Game Meat Pty Ltd (http://www.sgm.com.au/Species2.htm)
of inland Australia in areas receiving less than 500 mm mean annual rainfall (Caughley et al. 1987). Eastern Grey kangaroos generally inhabit areas of higher rainfall between the inland plains and the coast. They have an almost continuous distribution down the eastern seaboard where annual rainfall exceeds 250 mm. This distribution includes all of Queensland (except western Cape York), New South Wales, Victoria and north-eastern Tasmania (Caughley et al. 1987). The Western Grey kangaroo occurs across the south of the continent with a distribution extending northwards through western New South Wales and into a small area of southern central Queensland (Caughley et al. 1987, Strahan 1991).

Kangaroos are capable of breeding in all months of the year. Above average temperatures and rainfall stimulate the breeding cycle (Caughley et al. 1987). Breeding commences during the onset of monsoonal conditions in summer in northern Australia, and with late winter/spring rains in southern and eastern Australia. Breeding peaks with maximum plant growth, which for the Eastern Grey kangaroo is in summer (Strahan 1991). The Western Grey kangaroo has a strong peak of births between September and December (austral spring: Caughley 1987). On average, rain falls uniformly across seasons and sporadically in arid and sub-arid zones, and is unpredictable across a year and from one year to the next (Robertson et al. 1987). Accordingly, the birth rate in Red kangaroos is uniform across the year under average to good seasonal conditions (Shepherd 1987). The dynamics of kangaroo populations have not been observed to differ with population density (Bayliss 1987).

The birth interval for kangaroos is around eleven months (Shepherd 1987, Strahan 1991), and the juvenile begins to leave the pouch (i.e., become potentially exposed to infected mosquitoes) at around nine months in times of plentiful food supply (Caughley et al. 1984, Strahan 1991). Bayliss (1987) observed that kangaroos born just after a drought were not detected by aerial survey until twelve to fifteen months later.

1.3.3.3 Vertebrate immunity

The relationship between mosquito density and infection rates in a region is not always linear. Mosquito numbers in themselves cannot be considered a sufficient indicator of RRv disease, because although large numbers of human cases have been observed concurrent with large mosquito populations (for example Marshall et al. 1982, Dhileepan
1996, Ryan et al. 1999), excessive mosquito numbers do not always result in an outbreak of disease (Whelan et al. 1996, Lindsay and Mackenzie 1997). The likely explanation for this paradox lies with the moderating influence of vertebrate host immunity on the circulation of virus in the natural cycle. To achieve a high level of circulating virus with consequent spill-over of infection into the human population, a highly susceptible host population is required. If a high proportion of the host population is already immune, then virus amplification will not occur in the presence of abundant mosquito populations.

Following a viraemic phase, vertebrate host species develop an immunity to Ross River virus (Kay and Aaskov 1989). The period of viraemia appears to be short, and has been reported to last 0-5 days in several wallaby species, 6 days in Eastern Grey kangaroos, and 1-6 days in bandicoots and mice (reviewed in Kay and Aaskov 1989). The difference in the duration of the viraemic period between juveniles and adults, if any, is not known.

In a typical open population, in- and out-migration and birth and death provide a changing population-at-risk over time. If replenishment of the susceptible host population is fast enough compared with the dynamics of the pathogen, then the pathogen can persist and become enzootic (Halloran 1998).

The level of herd immunity in the vertebrate host population is principally a function of the (i) juvenile recruitment rate, and the (ii) vertebrate infection rate. Regarding the first point, the typical kangaroo breeding cycle indicates that recruitment of a susceptible population of juveniles does not occur in the same year that breeding commences, and that a lag of about one year (i.e., from one spring or summer to the next) is generally needed for a juvenile cohort to become exposed. Regarding the second point, environmental conditions that support large numbers of mosquitoes would be expected to trigger circulation of virus in the natural cycle (i.e., an exchange of infection between vector and host). Under these circumstances, the proportion of non-immune vertebrates in a population would decrease, and would remain low until the population was restocked with susceptibles. Therefore, both the time since the last outbreak and the occurrence of suitable environmental conditions in the year prior to the outbreak year are likely to be critical factors in determining the de-facto level of herd immunity in a population, and hence the risk of RRv disease outbreaks.
1.4 ROSS RIVER VIRUS IN HUMANS

1.4.1 Early understandings of the disease

The first published description of what is now called RRv disease is usually attributed to Nimmo (1928), who reported an outbreak of at least 100 cases near Narrandera, New South Wales, in the autumn of 1928. The main symptom described was a painful swelling of the joints. In the same year, Edwards (1928) reported an outbreak of “atypical dengue fever” in Hay, New South Wales, which had started six weeks before the Narrandera epidemic. In retrospect, the similarities between the two outbreaks make it likely that the second one was also a record of epidemic polyarthritis cases (Marshall and Miles 1984). In 1886, 30 people in western Victoria were reported as having experienced symptoms of fever (lasting two to three days), and a rash and severe pain that lasted “much longer” (Weber 1886). A recent reappraisal of that report suggests – through a process of elimination – that the symptoms described by Weber were consistent with a diagnosis of RRv disease (Wolstenholme 1993).

During World War II, troops were stationed throughout northern Australia and several outbreaks of epidemic polyarthritis were reported, coincident with the timing of the wet season in that region. Halliday and Horan (1943) first described a disease (which they named “acute polyarthritis”) in troops stationed between Darwin, Adelaide River and Larrimah. More than 100 cases were admitted to two army hospitals between November 1942 and January 1943, although “at least twice this many cases were observed elsewhere”, and similar cases had been reported in the months preceding November. Cases reported in the latter stages of the epidemic were “much milder and the full syndrome was rarely seen”. Culicine mosquitoes were observed in the region at that time.

Sibree (1944) reported 28 cases of “acute polyarthritis” in late February to early March 1944 in a military camp in northern Queensland (region unspecified). The possibility of the epidemic spread of the disease, combined with symptoms of persistent joint pain, rash, and adenitis in the patients, suggested the possibility of a diagnosis of several viral diseases such as dengue, rubella, and varicella.

Dowling (1946) observed 94 cases of polyarthritis among troops in northern Queensland from February to April 1945 (the third successive year during which such an outbreak
Chapter 1 - Research context

had occurred in the normal wet season in northern Australia). The outbreak commenced after two weeks of heavy falls of rain, which turned the region “into a swamp”. At the time the first cases were identified, Dowling noticed that mosquitoes had “become much more numerous since the heavy rain”. Anti-mosquito measures were introduced (including draining and oiling the swamps, insecticiding the tents and the use of mosquito nets at night). The preventive measures began one week after the first cases appeared, and the incidence of new cases declined rapidly after two weeks (about seven days after mosquito control began). Dowling renamed the disease “epidemic polyarthritis” rather than acute polyarthritis, on the basis that joint involvement was usually mild and was sometimes prolonged, and “to avoid confusion with other more serious forms of polyarthritis” (Dowling 1946).

These reports from the military outbreaks of the 1940s strengthened a hypothesis that insects spread the disease. The restricted movement and closely grouped conditions of troop populations, the common environmental exposures of the majority of cases, as well as the presence of trained entomologists to observe indigent insect life helped refine earlier speculations linking outbreaks to an infectious agent transmitted by arthropods or non-human vertebrate hosts (Anderson and French 1957). Stinging flies, March flies, and mosquitoes were all mentioned as possible vectors in these outbreaks.

An outbreak of epidemic polyarthritis was reported on the Atherton Tablelands of Queensland in 1949 (Short 1949), and then several outbreaks were reported from the Murray Valley region of south-eastern Australia in 1956 and 1957. Anderson and French (1957) described a “widespread epidemic”, with an estimated 1000 to 2000 cases between April and mid-May 1956. Symptoms were listed as “mild”, and included rash, joint and muscle pain. An absence of fever was noted. Cases of disease were limited to those people who resided within the Murray Valley region, or who had “been in close association with the [Murray] river” one to two weeks prior to infection (Anderson and French 1957). In attempting to explain the etiological aspects of the 1956 epidemic, Fuller and Warner (1957) noted that more mosquitoes were present in the region than usual, and that several thousand susceptible people had gathered for a regatta on Lake Bonney in Barmera at the beginning of April. People from all the towns in the region involved in the epidemic attended the event, and the first cases were recognised a few days after its conclusion. Such a gathering, they hypothesised, could account for a sudden outbreak of disease, and would support a theory that the causal agent was a virus transmitted by mosquitoes. They, and other authors reporting on the same outbreak, concluded that the
cases reported were sufficiently similar to be labelled epidemic polyarthritis, and that the cause was probably an arthropod-borne virus (Anderson and French 1957, Fuller and Warner 1957, Wilson 1957, Shope and Anderson 1960).

It was not until 1963 that the causative agent of epidemic polyarthritis was identified as an alphavirus, transmitted by a mosquito vector. In 1960, Shope and Anderson found that patients with epidemic polyarthritis developed antibodies to two other Group A arboviruses - Sindbis and Getah. They speculated that the causative agent was similar to, but distinct from, these two viruses. This view was confirmed by Doherty and others (1963), when they isolated RRv from *Ae. vigilax* mosquitoes collected near Ross River at Townsville in early 1959.

Ross River virus was first isolated from a human in 1972 (Doherty et al. 1972), although the symptoms of the case (an Aboriginal child from Queensland) were not considered typical for a diagnosis of epidemic polyarthritis. It was not until 1981 that the virus was isolated from a person on mainland Australia with recognisable symptoms of epidemic polyarthritis (Aaskov et al. 1981a). This isolation established RRv as a causative agent of epidemic polyarthritis.

Since then, it has been discovered that Barmah Forest virus is also capable of causing an epidemic polyarthritis-like illness (Mackenzie and Smith 1996). The close similarity of clinical symptoms between Ross River virus disease and Barmah Forest virus disease means that a laboratory confirmed diagnosis is the only certain method for distinguishing between them.

There is always the possibility that these early descriptions of epidemic polyarthritis, prior to the availability of a test for Ross River virus, may represent infection caused by other arboviruses (such as Barmah Forest or dengue). Dengue fever is "protean in nature" and has characteristics in common with epidemic polyarthritis (including afebrile or mildly febrile symptoms, joint pains which appear as periarthritis, and a rash that can be described as maculo-papular) (Sibree 1944). In addition, rubella, mild acute rheumatic fever, or mild rheumatoid arthritis may also exhibit similar symptoms to epidemic polyarthritis (Clarke et al. 1973). Paired sera from nine cases of RRv disease from 1968 and 1971 that were suspected, but serologically unconfirmed at the time, were retrospectively investigated for diagnostic rises in rubella. Two patients (22%) tested seropositive for rubella (Clarke et al. 1973). In summary, the early descriptions of
epidemic polyarthritis can only be suggestive of RRv activity in a particular region at a particular time.

There were only seven RRv disease outbreaks documented in the 80 years before a test for the causative agent was diagnosed (1886 to 1956/57). It is reasonable to assume that this record presents an incomplete picture of the real periodicity of disease in the regions (and of RRv activity elsewhere). Since that time, improved diagnostic techniques, an increased awareness of the disease, and the growth of human populations and their expansion (via permanent settlement and tourism) into natural RRv habitat, are all likely to have affected the increase in reporting of RRv disease outbreaks that has been observed.

1.4.2 Illness and incapacity

1.4.2.1 Symptoms

RRv causes a syndrome of disease symptoms. When it was determined that both RRv and BFv can cause epidemic polyarthritis, the syndrome was for a time referred to as “Ross River fever”. It is now commonly called Ross River virus disease, in recognition of the fact that not all patients present with fever or joint pain (Flexman et al. 1998).

An extensive review of the symptoms of RRv disease has been compiled by several authors (Mudge and Aaskov 1983, Fraser 1986, Fraser and Marshall 1989). Mudge and Aaskov (1983) reported that the virus caused joint symptoms, typically in the ankles, knees, finger, and wrists, in about 80% of cases, and myalgia in 58%. Fraser and Marshall (1989) recorded other main symptoms as a maculo-papular rash in 50% of cases, fever in 30-50%, and fatigue in some 50%. A wide diversity of symptoms has been reported during outbreaks, with some suggestion that differences in severity, duration, and type of symptoms can occur between geographic regions both within Australia and elsewhere (Condon and Rouse 1994).

Although the disease is not fatal, or permanently disabling, RRv disease can cause considerable pain, distress and loss of productivity. Ryan and others (1999) recently noted that of the 94 communicable diseases for which data were collected in the State of Queensland (1992 to 1996), RRv disease was the greatest cause of morbidity in 3 out of the 5 years, and was arguably one of the greatest ongoing health problems in that state.
1.4.2.2  Duration

The extent of long-term incapacity is still a matter of discussion. Harley and others (2001) reviewed the history of the duration of RRv disease. They noted an increase in the reported length of duration of symptoms since RRv disease was first described – initially from three weeks (Nimmo 1928, Halliday and Horan 1943) to three months (Dowling 1946). In the 1980s and 1990s, three studies provided estimates that indicated much higher rates of morbidity, and extended periods of incapacity. In these studies, more than 50% of cases reported that symptoms of arthralgia persisted after 12 months (Condon and Rouse 1994, Selden and Cameron 1996, Westley-Wise et al. 1996). Most of these studies were retrospective, however, did not use objective measures of ill-health, or adequately differentiate between RRv disease and other confounding diseases (Mylonas et al. 2001). There is also the possibility that, as most diagnoses in these studies were presumptive (i.e., not serologically confirmed), up to 20% of cases may have been mis-diagnosed (Lloyd et al. 2001a).

Recent prospective studies suggest that progressive resolution of symptoms in cases of RRv disease over three to six months is usual, and that incapacity in patients past six months may have been overestimated and due to pre-existing chronic rheumatic conditions (Mylonas et al. 2001, Harley et al. 2002). Despite some methodological limitations of these two studies (including moderate sample sizes, possible recruitment bias, and the absence of a control group), the biases are likely to be less than in the previous retrospective studies.

1.4.2.3  Treatment and prevention

There is no specific treatment for the disease. Condon and Rouse (1994) reported that one third of patients (n=255) found non-steroidal anti-inflammatory drugs to be most helpful, 25% preferred rest, 16% aspirin, and 10% physical therapies (swimming, massage, etc.). In a study by Mylonas and others (2002), a higher percentage of patients used non-steroidal anti-inflammatory drugs (58%, n=67).

Although RRv disease vaccination is not currently available, research is being conducted into the development of a vaccine (Yu and Aaskov 1994, Aaskov et al. 1997) and human trials are being considered. In the absence of a vaccine, prevention remains the sole public health strategy. The current approaches to disease prevention are discussed in Chapter 8.
1.4.3 Incubation period

Estimates of the incubation period for RRv disease (time from date of infection to date of onset of symptoms) vary at the extremes (Table 1.4). However, among those surveys that have been reliably conducted, there has been a consistent finding of a mean period of seven to eleven days. The strongest evidence comes from Fraser and Cunningham (1980). They personally interviewed 20 sero-confirmed cases who were not residents of an RRv enzootic region, and who contracted the infection while on holiday. They noted that if the results from their research were combined with those from a survey conducted by Anderson and French (1957), then 7 to 11 days would cover all but 1 of the 34 patients in both series.

<table>
<thead>
<tr>
<th>Author</th>
<th>Incubation Period, days</th>
<th>Sample size</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson &amp; French, 1957</td>
<td>10-11</td>
<td>14</td>
<td>Retrospective study based on travel histories; assumption that infection followed contact with Murray River.</td>
</tr>
<tr>
<td>Fuller &amp; Warner, 1957</td>
<td>3-15</td>
<td>5</td>
<td>Same as above.</td>
</tr>
<tr>
<td>Mudge &amp; Aaskov, 1980</td>
<td>9*</td>
<td>528</td>
<td>Cases not necessarily from a non-RRv enzootic area; self-administered questionnaire (51% response rate); sero-status confirmed.</td>
</tr>
<tr>
<td></td>
<td>(range 3-21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraser &amp; Cunningham, 1980</td>
<td>7-9</td>
<td>20</td>
<td>Cases from a non-RRv enzootic area; interviewer-administered questionnaire; sero-status confirmed.</td>
</tr>
<tr>
<td></td>
<td>(7-11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosen et al. 1981</td>
<td>3-5</td>
<td>2</td>
<td>Cases of infection in the Cook Islands; cases from a non-RRv enzootic area; sero-status confirmed.</td>
</tr>
<tr>
<td>Weinstein &amp; Cameron, 1991</td>
<td>12</td>
<td>11</td>
<td>Cases from a non-RRv enzootic area; interviewer-administered questionnaire.</td>
</tr>
<tr>
<td></td>
<td>(range 5-19)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Population mean.

The possibility of incubation periods as short as three and five days have been reliably reported, based on two serologically confirmed cases in the Cook Islands (Rosen et al. 1981). A theoretical maximum period of three weeks was noted in a study conducted by Mudge and others (1980), although this came from a self-administered questionnaire response from a person who may have resided in an RRv enzootic area, and cannot be considered definite. The current estimation of RRv disease incubation used by the Greater Murray Health Area (in southern New South Wales) is between eleven and fourteen days (Terence Carvan pers. comm. 2000).
1.4.4 Incidence and prevalence of RRv infection

There is no evidence of an RRv disease patient having a second clinical infection (Kay and Aaskov 1989, Kapeleris et al. 1996). The current understanding is that a person once infected with RRv will have sufficient antibody to prevent re-infection after subsequent exposure (Aaskov et al. 1997).

The incidence of clinical infection (ratio of sub-clinical to clinical infections) has been observed to vary dramatically. One study in Queensland reported a ratio of 80:1 (Kay and Aaskov 1989), a second from Fiji recorded ratios of between 4:1 and 2:1 (Aaskov et al. 1981b). In a large outbreak in New South Wales in 1983-84, Hawkes and others (1985) reported a ratio of 1.3:1 (340 RRv infections and 257 cases). And a report from an outbreak amongst troops in Queensland in 1997 (Russell et al. 1998) found a ratio of 1.2:1. In a recent review of the methods and assumptions underlying these published estimates, Harley and others (2001) concluded that the most reliable estimate to date of the clinical to sub-clinical infection ratio lies between 3:1 and 1.2:1. They also note that Kay and Aaskov’s (1989) theory - that the marked difference in ratios may relate to differences in population immunity levels between endemic areas and epidemic areas - cannot be based on sound evidence (Harley et al. 2001).

Antibody prevalence surveys (both prospective and retrospective) have been conducted to examine population rates of infection of RRv. Harley and others (2001) have comprehensively reviewed the topic. Several of their points are worth summarising. First, RRv infection clearly varies by geographic region. The sero-prevalence rates from the eleven studies conducted thus far strongly suggest a gradient from northern (warmer) to southern (cooler) regions, with rates higher in the north and lower in the south. Second, RRv antibody prevalence in most studies appears to increase with age. This contrasts with the observed rates of RRv disease (discussed below), which are highest in the middle age groups. It should be noted, however, that for most of the studies there was no information to determine whether the age structure of the denominator populations was the same as for the sample populations. Third, there was an apparent sex bias in most prevalence studies towards males, who had rates that were consistently a little higher than for females. Whether this was due to overrepresentation of males in the study population or not clear, although in at least one study there is evidence that more males were sampled than females (Weinstein et al. 1994).
A number of authors have observed that children appear to have a higher ratio of sub-clinical to clinical infection than adults (Fuller and Warner 1957, Wilson 1957, Seglenieks and Moore 1974, Mudge 1977, Mudge and Aaskov 1983, Hawkes et al. 1985, Kay and Aaskov 1989, Tai et al. 1993). This is supported by the results of a number of seroprevalence surveys, which have reported RRv antibody rates in children of as high as 45% (Hawkes et al. 1993) and 31% (Clarke et al. 1973).

1.4.5 Geographic and temporal distribution

1.4.5.1 Within Australia

Chapter 4 is a description of the epidemiology of RRv disease in Australia, based on notifications made to the National Notifiable Diseases Surveillance System (Communicable Disease Network) for the period July 1991 to June 1999. In the following sub-section, the major points relating to temporal and spatial distribution are briefly noted.

Outbreaks of RRv disease have been reported in all States and Territories of Australia. The epidemiology of the disease has been noted to vary across Australia (Russell 1994), reflecting the multiple vector and host species implicated in transmission, and the impact of diverse climatic and environmental conditions on their biological processes. In addition to the early reports mentioned previously, outbreaks have been reported in New South Wales (Clarke et al. 1973, Hawkes et al. 1985, Hueston and Cunningham 1989, Westley-Wise et al. 1996, Amin et al. 1998, Brokenshire et al. 2000), Victoria (Campbell et al. 1989, Wolstenholme 1991), Tasmania (McManus and Marshall 1986), South Australia (Seglenieks and Moore 1974, Mudge 1977, Mudge et al. 1980, Weinstein and Cameron 1991, Russell 1994), Western Australia (Lindsay et al. 1992, Lindsay et al. 1995), the Northern Territory (Merianos et al. 1992, Tai et al. 1993, Russell 1994), and Queensland (Ritchie et al. 1997).

In Western Australia, the greatest number of RRv cases are generally recorded in late spring and summer (Mackenzie et al. 1994a). In south-eastern Australia, cases occur in the late summer to autumn months (Kay and Aaskov 1989). In the tropical and sub-tropical northern and north-eastern regions, although cases of RRv occur all year round, the highest incidence is from mid-summer to autumn (Marshall 1979). The unusually broad range of known (and potential) mosquito vectors for RRv (Marshall 1979), coupled with
the different seasonal timing of rainfall, is likely to provide an explanation for the
different timing of RRv activity around the country.

1.4.5.2 Outside Australia

RRv disease has been reported in several countries in the region surrounding Australia,
including unconfirmed cases from the Schouten Islands (north of West Papua) in 1944-45
(Weber et al. 1946), Papua New Guinea (Scrimgeour et al. 1987), and a very large outbreak
across the Pacific in 1979-80. In the latter outbreak, cases were serologically confirmed in
Fiji (Aaskov et al. 1981b), and serological studies indicated there were six to ten sub-
clinical infections for every infection – which represented some 300,000 to 500,000
infections across the Fiji islands (Marshall and Miles 1984). This outbreak spread to
American Samoa (Tesh et al. 1981), and later to the countries of New Caledonia and the

New Zealand is at risk of the establishment of RRv due to the recent introduction of a
highly competent vector, *Ae. camptorhynchus*, that has been found on several occasions
and at multiple locations in New Zealand’s north island since December 1998 (Robinson
2002). The main reservoir hosts do not occur widely in New Zealand, although
substantial wallaby populations exist in discrete parts of the North and South Islands,
and are also found near the major centre of Auckland (Graeme 2002). Even so, the
country is expected to be vulnerable to RRv transmission due to the extremely large
urban and rural feral possum population (Kelly-Hope et al. 2002). In addition, the risk is
increased by the frequent travel of infected people from Australia to New Zealand, and a
non-immune resident population (Knol et al. (In Press)).

1.4.5.3 A rural disease?

RRv disease has been previously referred to as a “disease of rural Australia” (for example
Hawkes et al. 1985, Merianos et al. 1992), reflecting the location of the historical outbreaks
of disease. However, a number of outbreaks in the 1990s have demonstrated the potential
for RRv activity to spread from semi-rural areas into the outskirts of major metropolitan
centres. In early 1997, 69 cases of RRv disease were reported in the north-western suburbs
of Sydney, a rural setting in the Sydney basin. Subsequent investigations revealed that
some 71% may have been infected locally (Amin et al. 1998). In early 1999, 25 cases were
confirmed in outer-western Sydney (Brokenshire et al. 2000), of which 17 were later
believed to have been infected inside the metropolitan area. This region is close to natural
woodlands, with abundant macropod hosts and local vector breeding. Outbreaks have also been recorded in the temperate city of Perth (Lindsay et al. 1992, Lindsay et al. 1996a), and in Brisbane (a sub-tropical city on the eastern seaboard: Ritchie et al. 1997).

Dale and others (1998) analysed the incidence of RRv disease in Brisbane by street address and found a statistically significant difference in the distribution of RRv disease notifications. The spatial pattern to disease incidence pre-supposes that some of the annual notifications in Brisbane can be related to local, rather than imported, infection (they did not attempt to quantify this proportion). On a sample area of the city, they hypothesised that risk increased with residential proximity to freshwater breeding sites (such as depressions that hold water for up to a week). No doubt, greater effort will be put into researching this matter in the future, as residential development into mosquito habitat is increasing in all major cities in Australia.

1.4.6 Factors moderating RRv transmission and infection

1.4.6.1 Age, sex, and race

Children and older people typically record lower rates of disease than adults aged 20 to 55 (for a review see Kay and Aaskov 1989, Harley et al. 2001). The reasons for this may be three-fold. First, children appear to have a different immune response to infection, and show a lower rate of clinical expression (discussed previously). Second, both children and older people are likely to have a lower exposure to mosquitoes than people in their more active middle years. Third, older people are more likely to have become immune through previous infection (Boughton et al. 1984).

Studies have shown conflicting evidence for the ratio of disease in males compared to females. Mudge (1983) reported the ratio to be 0.6:1, and Hawkes (1985) reported a similar ratio of 0.8:1. Since then, other studies have reported ratios of 1.7:1 (Westley-Wise et al. 1996), 2:1 and 0.8:1 (Aaskov et al. 1981a) and, varying by age category, 0.7:1 and 1.25:1 (Selden and Cameron 1996). These reported differences may relate to greater occupational or recreational exposure to vectors (Kay and Aaskov 1989), or to selection or response bias in the study populations. There is no evidence (or suggestion) that RRv infection provokes a greater immunologic response in one sex than the other.

Doherty and others (1966) noted that Aboriginal communities in Queensland had higher sero-prevalence rates than non-Aboriginal communities. During an outbreak in the
Northern Territory in 1990-91, Tai and others (1993) noted a much lower disease incidence amongst Aboriginal than non-Aboriginal people. They speculated this was due to the higher exposure of rural Aboriginal people in early life to mosquitoes (due to proximity to bushland, unscreened houses, etc.) and subsequent asymptomatic seroconversion and immunity. There is no evidence of different infection rates among other races.

1.4.6.2 Human behaviour and lifestyle

A few studies have demonstrated the effectiveness of a range of individual behaviours in reducing the risk of infection with RRv (Weinstein and Cameron 1991, Westley-Wise et al. 1996). Such behaviours include the wearing of long-sleeved loose clothing, the use of personal mosquito repellent, and the removal of domestic breeding sites. The protective effect of insect proofing buildings and water tanks against mosquito bites has also been observed in relation to dengue infection risk (Murray-Smith et al. 1996). Observations from RRv disease outbreaks have noted a higher incidence of cases amongst some occupational groups (for example, housewives in Mudge et al. 1980, and people working outdoors in Hawkes et al. 1985), although these were not necessarily drawn from representative population samples.

1.4.6.3 Land use changes

Peri-urban sprawl is a persistent feature of Australian population re-distribution, and is predicted to continue (State of the Environment Advisory Council 1996a). As well as intensifying pressures on coastal systems, it also amplifies the contact of non-immune populations with the virus' natural cycle. Several studies have noted the increased risk of RRv infection in regions where residences are closely located to wetlands and bushland (Lindsay et al. 1996a), and freshwater swamps (Whelan et al. 1996). Although there is an awareness of the need for town planning mechanisms to consider the appropriate siting of human habitations in most RRv enzootic areas (i.e., with reference to proximity to mosquito breeding areas), only a few localities have made this mandatory in Australia. The growth of tourism into northern tropical and sub-tropical Australia, predominantly to non-urban areas, is also likely to expose increasing numbers of overseas visitors and local inhabitants to the risk of RRv disease (Tourism Forecasting Council 2002).

Irrigation methods for some crops (discussed in the next section) also provide impermanent breeding sites for mosquitoes in some regions (Russell 1994), thus
extending the annual transmission season. Australia is a dry country with limited water resources. Despite this, it has one of the highest total water consumption levels per head by international standards. Constructed wetlands for wastewater treatment are being encouraged in Australia as a way of improving effluent water quality and reducing total water consumption (State of the Environment Advisory Council 1996b). Engineering and water quality objectives in this matter may conflict with mosquito management interests (Russell 2001). Research is being conducted to determine the characteristics of wetlands that will facilitate the emergence of mosquitoes under different climatic conditions in Australia (Dale et al. 2001).

1.5 **WEATHER, CLIMATIC AND ENVIRONMENTAL INFLUENCES ON THE ARBOVIRAL CYCLE**

The transmission dynamics of arboviral diseases are dependent on many complex and interacting factors that determine whether outbreaks occur, and if patterns of disease will be endemic or epidemic. These include human population density and immunity; housing location; the availability of screens and air conditioning in houses; human behaviour; the availability of piped water; land use and irrigation systems; the availability and effectiveness of mosquito control programs; as well as sufficient numbers of susceptible reservoir hosts, mosquito vectors and pathogen (Gubler 1998b, IPCC 2001a). In addition, the biology and distribution of the vector, the replication of the virus in the vector, and the breeding of vertebrate hosts are profoundly influenced by both climate and weather (Sellers 1980, Edman and Spielman 1988, Reiter 1988, Kay and Aaskov 1989, Russell 1998b). When the aforementioned factors are favourable, then weather and climate factors can play a dominant role in influencing the transmission intensity and distribution of mosquito-borne diseases (McMichael 2001).

1.5.1 **The weather-climate distinction**

The term "weather" refers to the "state of the atmosphere in a particular place and time, with respect to wind, temperature, cloud, moisture, pressure, etc (Macquarie Library 1989). For example, the maximum temperature on a given day in Canberra might be 20°C. Weather data are transformed into climate data if summarised over time. "Climate" refers to the weather prevailing in an area over a long period. For example, the average maximum summer temperature in Canberra is 26°C. For the purposes of this thesis, two types of climate data can be distinguished.
Chapter 1 - Research context

- First, daily weather measurements, averaged for a discrete period (e.g., January rainfall, Canberra, 2003), represent the climate of a region based on that particular period of observations. The daily weather measurements used in this thesis were averaged into monthly climate values (see Chapter 3).

- Second, long-term climate (i.e., seasonal, annual, inter-annual, decadal, or longer time-frames) represents the average weather experience in a region, based on many years of observation. The long-term climate averages used in this thesis were based on 50 years of observations (also in Chapter 3). In terms of climate change, the “baseline” climate against which change is measured is taken from averages for the period 1961-1990 (IPCC 2001b).

In relation to the subject of this study, Epstein (1998) has proposed “it is axiomatic that climate circumscribes the range of vector-borne diseases, whereas weather affects the timing and intensity of outbreaks”. The following sections describe the biology underpinning this relationship, and refer to epidemiological studies that have recognised, or tested, the relationship between weather/climate and outbreaks of mosquito-borne diseases, with particular focus on research relating to RRv disease.

1.5.2 Rainfall

Water is essential for the breeding cycle of mosquitoes, as the larval and pupal stages are aquatic (Sellers 1980). There is substantial evidence associating heavy rainfall and, by extension the breeding of mosquitoes, with outbreaks of RRv disease. In temperate southeastern Australia, abnormally heavy rainfall or extensive flooding has been reported preceding, or contemporaneous with, the majority of outbreaks of RRv disease (Nimmo 1928, Anderson and French 1957, Fuller and Warner 1957, Wilson 1957, Seglenieks and Moore 1974, Mudge and Aaskov 1983, Hawkes et al. 1985, Finn 1995) and of Murray Valley encephalitis (Forbes 1978, Nicholls 1993). In most districts of Western Australia, the principal indicator of increased risk of RRv transmission in the period 1984 to 1993 was heavy rainfall just prior to or during the established season of risk (Lindsay 1995, Lindsay et al. 1996a). In tropical parts of the country, outbreaks of RRv disease have tended to coincide with the wet season (Halliday and Horan 1943, Sibree 1944, Dowling 1946, Whelan et al. 1996). An outbreak in the lower latitudes of the Northern Territory in 1991, for example, was attributed to the highest rainfall recorded in the preceding decade, accompanied by optimal humidity and temperature for mosquito breeding, abundant reservoir populations, and human population in-migration (Tai et al. 1993).
An association with periods of heavy rainfall is evident in most mosquito-borne disease outbreaks, but the effect of rainfall on mosquito breeding is not always a direct and positive one. A moderate increase in rainfall can be beneficial (Lindsay and Mackenzie 1997), although excessive increases can wash away the mosquito larvae or dormant eggs and interrupt the transmission cycle, a particular problem for species that prefer to breed in still water (Paul Reiter, interviewed in Taubes 1997). In addition, insufficient susceptible vertebrate hosts can mean that heavy flooding will not lead to an outbreak of disease (Mackenzie et al. 2000). In a given period, the timing of an “event” (i.e., the number of contiguous days with recorded rainfall), the number and duration of events, as well as the total precipitation (i.e., amount) of an event, are factors that can differentially affect mosquito breeding. Thus, all must be considered as separate indicators of arbovirus disease outbreaks, and are of critical importance in particular in epizootic regions (Davies et al. 1985).

The timing of rainfall, and in particular the seasonality (Mackenzie et al. 2000), is decisive in determining whether an outbreak will occur, and if it does whether it will be epidemic in proportion. RRv disease outbreaks in the arid zones of Western Australia between 1984 and 1993 were more likely to appear following heavy autumn or winter rains, but did not occur during heavy summer rains (Lindsay and Mackenzie 1997). Winter or spring rainfall in arid regions is necessary to allow time for RRv to amplify to sufficient levels in the vertebrate population. In the temperate south-west, however, above average late-spring to summer rains were positively correlated with the largest outbreak of RRv disease recorded in that region (Lindsay et al. 1996b). These conditions were conducive to supporting the widespread breeding and survival of the major regional RRv vector, Ae. camptoryynchus, into warmer summer months (Lindsay et al. 1996b).

The number of rainfall events in a period has been observed to act as a crude surrogate for relative humidity levels in outbreaks of several arboviral disease. Davies and others (1985) found only a low correlation between Rift Valley fever epizootics and rainfall amount over a 32 year period. However, they found that a composite rainfall statistic that incorporated the persistence of rainfall at each station, as well as the monthly mean, was well correlated with epizootics (Davies et al. 1985). During an outbreak of RRv disease in 1994 around Brisbane, cases were notified from January to June (austral summer to beginning of winter), with the peak in March and April (austral autumn) (Ritchie et al. 1997). Mosquitoes were also observed to be abundant during March and April. Although total rainfall was below average in these peak months, the number of rain days was
elevated during February and March. Ritchie and others speculated that the high number of rain days led to higher than average humidity, and provided suitable conditions for the build-up of mosquito populations four to six weeks later. He proposed a composite mosquito breeding indicator statistic for RRv disease in tropical Australia, based on the number of rain days per month, rainfall amount, and temperature (Ritchie et al. 1997).

Rainfall also strongly affects the periodicity of the transmission cycle, by affecting the recruitment rate and survival of non-immune juvenile vertebrate hosts. This is particularly the case in epizootic regions. In arid and semi-arid regions (and in temperate regions in severe drought years), the main determinant of the rate of increase of kangaroo populations is the level of food supply (Bayliss 1987, Caughley 1987) - and rainfall is the dominant factor governing the control of pasture biomass (Noy-Meir 1973). Kangaroo populations respond to changes in food supply by adjusting their rates of reproduction, survival, or both (Bayliss 1987). In a study of the Menindee district in western New South Wales on the floodplain of the Darling River (a semi-arid zone), Caughley and others (1987) concluded that the instantaneous growth rate of pasture components was highly dynamic, that the growth of plants was best correlated with rainfall from either the previous one or two months, and that growth could be initiated by sufficient rainfall (i.e., more than 40 mm per month) in any season. Grasses in the Murchison area of Western Australia (an arid zone) have been reported to take only 24-27 days from a rainfall event to the initiation of flowering (Mott and McComb 1975). It has also been observed that Red kangaroos come into breeding condition almost immediately following rainfall and after drought conditions, and that if rainfall is high and food abundant, juvenile survival is high (Strahan 1991).

Depending on the breeding habits of local vector species (i.e., if they are wetland or puddle breeders), soil moisture may better explain mosquito breeding than weather factors such as precipitation and temperature - especially under conditions of extreme rainfall variability (Patz et al. 1998). Patz and others (1998) studied the relationship between modelled soil moisture and key factors for malaria transmission (i.e., mosquito biting rates and entomological inoculation rates). Modelled soil moisture (lagged by two weeks) proved to be a better predictor of biting rates than rainfall (correlation coefficient $r^2 = 0.45$), although rainfall combined with maximum and minimum temperature (lagged by three weeks) could account for almost the same amount of variability in the biting rate ($r^2 = 0.41$).
Hawkes and others (1985) noted that prolonged drought followed by heavy rainfall often occurs before extremely large outbreaks of RRv disease (such as those in 1983/84 in New South Wales). However, they also noted that this has not always been the pattern (i.e., the large RRv disease outbreaks of 1928 and 1956 did not follow drought periods). While extended drought periods may allow for the recruitment of susceptible vertebrate hosts, they are clearly not a pre-condition for an outbreak of RRv disease. As already noted, some number of years of low rainfall may be necessary to reduce viral activity in the natural cycle in epizootic regions.

1.5.3 Temperature and humidity

Temperature directly affects the distribution and nutritional requirements of mosquitoes, and the length and efficiency of the extrinsic incubation period of arboviruses in their vectors (Kay et al. 1989, Reeves et al. 1994, Mackenzie et al. 2000). The extrinsic incubation period is the time needed for a virus to replicate in the salivary gland of the mosquito to sufficiently high titres for infection to occur. As temperatures increase replication time reduces – thus arboviral transmission is enhanced because mosquitoes become infectious more quickly (Weinstein 1997). Table 1.5 summarises the effect of temperature on the breeding and virus replication time for the major RRv species.

### Table 1.5 The effect of temperature on breeding time and extrinsic incubation periods for the major Ross River virus species.

<table>
<thead>
<tr>
<th>Mosquito species</th>
<th>egg to adult days</th>
<th>temp. ºC</th>
<th>EIP range (lower to maximal)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Cx. annulirostris</em></td>
<td>12-13</td>
<td>25</td>
<td>7 ¹</td>
</tr>
<tr>
<td></td>
<td>9 ¹</td>
<td>30</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>12 ²</td>
<td>23-24</td>
<td>-</td>
</tr>
<tr>
<td><em>Ae. vigilax</em></td>
<td>6-8 ¹</td>
<td>25</td>
<td>10 ¹</td>
</tr>
<tr>
<td></td>
<td>&lt;6 ¹</td>
<td>30</td>
<td>-</td>
</tr>
<tr>
<td><em>Ae. camptorynchus</em></td>
<td>-</td>
<td>-</td>
<td>5-9 ³</td>
</tr>
<tr>
<td><em>Ae. notoscriptus</em></td>
<td>-</td>
<td>-</td>
<td>9-14 ⁴</td>
</tr>
</tbody>
</table>


Sudden temperature increases and decreases can result in reduced optimal vector breeding conditions. If warm weather and strong winds follow a period of heavy rain, then surface water will dry up, and the conditions will disappear unless more heavy rain quickly follows. In relation to RRv disease epidemics in south-eastern Australia, Hawkes (1985) speculated that, even in the presence of ideal rainfall conditions, a sudden period
of cold weather during summer was sufficient to delay the build-up of some species of mosquitoes and abort an epidemic.

Temperature can also lead to a reduction in the period needed for larval development, within the comfortable breeding temperature range of the mosquito (Mackenzie et al. 2000). For the RRv vector *Cx annulirostris*, ambient temperatures that persist in the region of 25°C are optimum for its full life cycle (Table 1.5). At higher temperatures, *Culex* larvae complete their development faster (a 5°C increase gives a 3-4 day egg to adult reduction), and more generations can fit into a given time period (Kay and Aaskov 1989).

Extreme temperatures will kill mosquito populations, and *Cx annulirostris* larvae have been observed to die at temperatures below 10°C and above 40°C (Lee et al. 1989). Larval development was optimised at 30°C (McDonald 1980), and survival was longest at 25°C (Lee et al. 1989). Temperatures above 35°C were detrimental to adult *Cx annulirostris* population growth (McDonald 1980).

Temperature and humidity together influence the survival rate of infected female mosquitoes (Reeves et al. 1994), which is a major contributor to the outbreak of arbovirus epidemics. In a computer-simulated model for the vector-host transmission cycle of a mosquito-borne avian virus, the increased survival of infected mosquitoes alone caused more infections than any other single component in the system (Scott et al. 1983). De Moor and Steffens (1970) found that mosquito survival was the most important factor in the chickungunya virus cycle they modelled. And Kay and others (1987) observed that mosquito survival, extrinsic incubation period, and duration of the mosquito feeding cycle - all affected by ambient temperature and humidity - were the most important variables influencing the speed of rural amplification of Murray Valley virus. The incidence of RRv disease in Toowoomba, Queensland, between 1986 and 1995 has been positively associated with minimum temperature and humidity (additional details not supplied: Tong 1998).

Relative humidity is the ratio of the amount of water vapour in the air compared to its total capacity at a given temperature. Relative humidity affects longevity, mating, dispersal, feeding behaviour and egg-laying of mosquito species (Mackenzie et al. 2000). High relative humidity increases the proportion of old mosquitoes in the population, thus increasing the risk of RRv transmission (as older female mosquitoes are more likely
than younger ones to have had two or more blood meals). Mosquitoes have high surface area to mass ratio and are susceptible to loss of body water if a rise in ambient temperature is not accompanied by a rise in humidity. This has particular implications for arid regions, and to a lesser extent for temperate ones. High rates of mosquito attrition were noted as possible reasons for an absence of RRv disease outbreaks during the hottest months in an arid region of Western Australia (Lindsay 1995).

Relative humidity can also act as a surrogate for temperature and rainfall, as it is high only when these other two weather factors are high. Hales and others (2002a) found that relative humidity was the single best predictor of the current global distribution of dengue fever (OR=1.3, SE=0.003, p<0.001) with 89% accuracy (Hales et al. 2002a). Humidity affects the evaporation of water, and can influence the availability of breeding sites in summer, a particular problem for vectors in hot, dry climates (Lindsay and Mackenzie 1997).

1.5.4 The El Niño-Southern Oscillation

1.5.4.1 Description

The prime example of inter-annual climatic variation in the Australian region is the El Niño-Southern Oscillation (ENSO). This phenomenon refers to an ocean-atmosphere interaction in the tropical Pacific, and to a lesser degree in the Indian Ocean. It has been described as the strongest source of natural variability in the global climate system (Philander 1990). ENSO influences regional precipitation through changes in the strength and location of the major convective centres, and through zonal teleconnections (Allan et al. 1996). The term “teleconnections” was popularised by Bjerknes (1966), who used it to describe the complex that moves mass and energy from the equatorial Pacific to other parts of the globe. In the context of this thesis, the term refers to the connections between ENSO events and distant weather or climate-related anomalies.

Allan and others (1996) provide a comprehensive overview of the ENSO phenomenon and its relationship to natural fluctuations in the climate system. Briefly, ENSO has two extremes – the El Niño and La Niña (terms derived from Peruvian culture where the phenomenon was first described). During an El Niño “warm event” the sea surface temperature (SST) in the eastern Pacific Ocean rises. This increase is accompanied by
changes in sea level atmospheric pressure across the Pacific Basin (Rasmussen and Carpenter 1982). When the pressure rises in the east, it usually falls in the west (around Australia), a “see-saw” effect observed by Sir Gilbert Walker in the 1920s. Accompanying an El Niño, changes in atmospheric circulation are evident as weaker westerly trade winds over the region of very warm water in the western Pacific. This variation in the Walker circulation is known as the Southern Oscillation (Hay et al. 1993). The Southern Oscillation Index (SOI) provides a measure of the strength and timing of ocean-atmosphere changes in the south-east Pacific Basin, and represents the pressure difference between Indo-Australasia (recorded at Darwin) and the south-eastern Pacific (Tahiti). Large negative values of the SOI indicate a warm water event (El Niño). Specific spatial patterns of climate fluctuations (droughts in continents such as Australia, heavy rains and floods in others) tend to occur because of these large scale ocean and atmospheric changes.

The term La Niña was introduced to describe the opposite phase of the cycle. During a La Niña, the pattern of climate anomalies is, typically, reversed (Nicholls and Katz 1991). Colder than average sea surface temperatures are recorded in the eastern equatorial Pacific (and warmer than average in the western Pacific, around Australia). Stronger than normal easterly trade winds occur throughout much of the tropical Pacific, in association with higher than normal atmospheric surface pressure at Tahiti and lower than normal pressure at Darwin. Large positive values of the SOI indicate a La Niña (cool water) event.

ENSO events tend to occur on average every four to five years (range two to seven years) (Trenberth and Hoar 1996). Estimates for the duration of an anomalously warm water period associated with an El Niño range from twelve to eighteen months (with “a few notable exceptions”: Glantz 1996), and “up to fourteen months” (Gill 1982). Once an El Niño develops it often (but not always) exhibits a quasi-biennial cycle that shifts from one phase to the other in something like an 18-24 month time span (Allan 1995).

In this thesis, the term “El Niño” is used to refer to anomalous sea surface temperature increases in the eastern Pacific (warm water events), and “La Niña” to anomalous sea surface temperature decreases (cool water events). I use “ENSO” to refer to the two oceanic components of the cycle, combined with the atmospheric component (the Southern Oscillation).
Chapter 1 – Research context

1.5.4.2 ENSO and mosquito-borne diseases: global teleconnections

The effects of the ENSO cycle on agriculture have been the subject of scientific interest since at least the 1972-73 El Niño event (Glantz et al. 1991, Nicholls 2002). Weather-related anomalies that had previously been viewed only from a local perspective, were understood to be linked because of the great impact of this event. Since then, evidence for the effect of the warm and cold cycles on global climate – particularly on rainfall levels, and therefore on crop failure and their economic consequences – has been systematically established (Macilwain 1995). Appreciation of the relationship between ENSO teleconnections and increases and decreases in human morbidity and mortality has been slower. For example, analysis of the agricultural and economic implications of El Niño in Peru preceded a recognition of its implications for health and health care by many years (Salazar-Lindo et al. 1997).

An increasing number of studies have demonstrated significant associations between El Niño teleconnections and mosquito-borne disease occurrence, distribution and severity. ENSO teleconnections can drive outbreaks of mosquito-borne diseases by a number of mechanisms. ENSO is a climate phenomenon that strongly influences weather phenomena such as cloud cover, run-off, the number of wet days, cyclone frequency, temperatures, humidity, and evaporation (Nicholls 1988a), with flow-on effects to local ecologies. For example, Hales and others (1996) found positive values of the Southern Oscillation Index (i.e., La Nina “wet conditions”) could be correlated with dengue outbreaks in the Pacific Islands (Spearman $r^2=0.58$, $p=0.002$) – a relationship that has also been observed in other tropical countries (Gagnon et al. 2001).

Changes in the natural habitats of reservoir hosts can increase their abundance, longevity, and range. In turn, there may be a proliferation of non-immune vectors, which transport the infectious agent to susceptible people, their food, or their immediate surroundings. In some cases, extreme weather fluctuations may cause either host or vector populations to migrate to areas outside of their habitat, in this way bringing them into contact with non-immune human populations. For example, the first recorded outbreak of hantavirus pulmonary syndrome in the United States was believed to be the result of a massive increase and expansion of the mouse population that occurred following extreme rainfall from the 1991-92 El Niño event (Glass et al. 2000).
Kovats and others (1999) examined the evidence (epidemiological, climatological, and biological) for establishing a causal relationship between ENSO and health. They concluded that "there is good evidence that the ENSO cycle is associated with the increased risk of certain diseases. The relationship is particularly strong for malaria, but is suggestive for other mosquito-borne and rodent-borne diseases" (Kovats et al. 1999).

Specifically, important associations between ENSO and mosquito-borne disease outbreaks have been identified for malaria (Bouma and van der Kaay 1994, Thomson et al. 1996, Kilian et al. 1999, Lindsay et al. 2000), dengue (Hales et al. 1996, Hales et al. 1999, Gagnon et al. 2001), Rift Valley fever (Linthicum et al. 1999), Murray Valley encephalitis (Nicholls 1986), and RRv disease (Tong et al. 1998, Maelzer et al. 1999).

1.5.4.3 ENSO teleconnections in Australia

The relationship between ENSO, climate and weather

Over Australasia, the effect of El Niño is to remove or weaken the major sources of moisture or cloudiness. McBride and Nicholls (1983) investigated seasonal relationships between the Southern Oscillation Index and rainfall in Australia, and identified a distinct seasonal cycle in correlations, with spring (September-November) the strongest, and summer (December-February) the weakest. The highest correlations were noted for parts of northern and eastern Australia (Figure 1.2). Areas of Australia likely to receive more than normal rainfall during a La Niña are the south-east, the central region, and the south-western tip (Figure 1.3). Tropical cyclone activity in eastern Australia, one of the several synoptic systems that affects rainfall, has been persuasively linked to ENSO cycles (Nicholls and Lavery 1992).

Nicholls has shown that areas strongly influenced by ENSO have higher relative rainfall variability than areas not affected (typically by one-third to one-half), and that low-latitude (i.e., tropical) areas show greater variability with ENSO events, therefore causing stronger drought and flood patterns (Nicholls 1988b, Nicholls 1998). Nicholls and Kariko (1993) examined the number, average length, and average intensity of rain events from 1910 to 1988, and found that ENSO affected rainfall mainly by influencing the number and intensity of rain events (as distinguished from rainfall totals).

El Niño-associated drought typically lasts about one year and breaks down in the late austral summer or autumn. On rare occasions, El Niño-like conditions persist through a
second year (Australian Bureau of Meteorology 2002b). The El Niño event of the first half of the 1990s lasted longer than any previous event in the instrumental record. This extended El Niño was accompanied by intense, persistent drought over central and southern Queensland, and northern New South Wales, extending at times to other parts of the Australian continent.

Impact of ENSO on RRv disease patterns

Severe outbreaks of RRv disease in temperate south-eastern Australia have generally followed heavy summer rainfall, which suggests an association with La Niña (Nicholls 1993). However, given the different ecologies of RRv across the country, and the inconsistent influence of ENSO on continental rainfall pattern, it is not surprising that Harley and Weinstein (1996) found no relationship between La Niña events and either RRv disease outbreak years or notified cases when analysed for Australia as a whole. Maelzer and others (1999) used logistic regression to model published reports of RRv disease outbreaks in the more geographically confined south-eastern Australian region (the States of New South Wales and Victoria) with the Southern Oscillation (for the year of the epidemic and lagged by one year). They found outbreaks of RRv disease were positively correlated with Southern Oscillation Index values. Tong and others (1998) compared Southern Oscillation Index values with the incidence of RRv disease in the town of Toowoomba, Queensland from 1986-95, and also found a positive correlation in that region.

1.5.4.4 The use of ENSO as an early warning indicator

An analysis of the Southern Oscillation Index and monthly rainfall in three locations in eastern Australia (1876-1997) found that the relationship between the two was contemporaneous (Cunningham and Poskitt 1998). Even so, it is possible to predict an El Niño (or La Niña) event some months ahead because the oceans have a persistent influence on weather systems. The “phase-locking” of the Southern Oscillation Index into an annual cycle (from early winter onwards) permits seasonal forecasting to be attempted once the early stages of various manifestations of the ENSO cycle are under way (Stone et al. 1996), albeit with caveats on the accuracy of predictions. A strong auto-correlation between Southern Oscillation Index values in early winter with those three to six months
Figure 1.2 Areas in Australia most likely to receive less than normal rainfall during an El Niño.


Figure 1.3 Areas in Australia most likely to receive more than normal rainfall during a La Niña.

ahead (i.e., phase-locking) has been noted (Gordon 1986), although the details of the accuracy of this for different regions of Australia is not available.

The longer lag time between the commencement of ENSO events and their effect on Australian rainfall or sea level, compared to the lag time of predictions based on "standard" weather variables (such as precipitation and temperature) (Nicholls 1988a), indicates the potential of ENSO indices (i.e., the Southern Oscillation Index and sea surface temperature) as early warning indicators of RRv disease outbreaks. September Southern Oscillation Index values were positively associated with RRv disease epidemics that typically commence in November or December in the south-eastern States of Australia (OR=1.23, 95% CI=1.08-1.39: Maelzer et al. 1999). Linthicum and others (1999) compared Rift Valley Fever activity in Kenya with monthly Southern Oscillation Index values, sea surface temperature values, and the NDVI for a 50 year period. They found that epidemics could be predicted up to five months in advance in East Africa.

The use of ENSO will be limited to areas where it strongly modulates the atmosphere or ocean (Nicholls 1988a). However, even in regions where ENSO appears to strongly modulate weather patterns, the use of ENSO as a predictor of mosquito-borne disease epidemics is complicated. El Niño events have variable frequency, magnitude, spatial characteristics, and duration (Philander 1990, Allan et al. 1996), each of which is likely to cause fluctuations in El Niño teleconnections. Quinn and others (1987) have ranked ENSO events on a scale from very strong to weak (accepting that, given the types of information upon which their historical records were based, intensity determinations were essentially subjective). Even in those parts of the globe most strongly and directly affected by ENSO teleconnections, there are vast differences in how a single event is experienced. The concept of a 'threshold effect' has been suggested by Nicholls (1998) to explain why there is sometimes poor correlation between the SOI and rainfall, except in extreme events. If true, it may be that only ENSO events of strong to very strong intensity will impact on the transmission of diseases (Nicholls 1998).

This variability of strength makes quantification of the impacts attributed to each event difficult. Should the prolonged El Niño which began in 1991 and ended in 1993 or early 1995 (depending on how it is interpreted) be considered one long event, or a few smaller ones (Glantz 1996)? Prediction based on ENSO necessarily involves uncertainty, due to the empirical base of the teleconnections as well as the theoretical limits of predictability inherent in the atmosphere-ocean system (Nicholls and Katz 1991). Research on
relationships between ENSO teleconnections and health impacts must currently work within these moving parameters and acknowledge the limitations of the science.

### 1.5.5 Tide height and sea-level

During a La Niña phenomenon, anomalous cooling of the sea water in the eastern Pacific (generally) corresponds to a warming of the sea water in the western Pacific, and an increase in the mean sea level – and hence high tide height – around northern and western Australia (Allan et al. 1996). Tidal inundation provides water for the breeding for RRv salt-marsh vectors, such as *Ae. camptorynchus*, *Ae. vigilax* (Lindsay and Mackenzie 1997) and *Ae. funereus* (Ritchie et al. 1997). Salt marshes and tidal flats along the temperate south-west coast of Western Australia have been inundated more frequently in years when Southern Oscillation Index values are neutral or positive (i.e., in a La Niña phase) (Lindsay et al. 1998), and positive values have been observed to favour salt-marsh mosquito breeding in that region (Lindsay et al. 1989).

Various descriptive epidemiological studies have noted an association between outbreaks of RRv disease and high tides in south-west Western Australia (Lindsay et al. 1993a, Lindsay 1995, Lindsay et al. 1996c, Lindsay et al. 1998), in Tasmania (McManus et al. 1994), in the Northern Territory (Whelan et al. 1996), and between incidence rates and tides in Queensland (Ritchie et al. 1997, Tong et al. 2002).

### 1.5.6 Environment

In this thesis, I use the term “environment” to mean the physical environment and its influences on human behaviour. I distinguish between “environment” and “climate”, which are sometimes used interchangeably.

#### 1.5.6.1 Vegetation

Vegetation is the manifestation of an area’s climatology, lithology, hydrology, topography, and human impacts. Thus the structure of vegetation tells something about the status of the environments they occupy (Carroll and Morain 1992). Vegetation has been used as a surrogate for ecosystems because it integrates other complex interacting environmental attributes (such as climate, soils, terrain and biota). Traditionally, ecologists have tended to use spatial classification of terrestrial vegetation as the dominant attribute defining ecosystems (Thackway and Creswell 1992).
Vegetation type and quantity influences the survival and breeding of RRv vertebrate host populations (Caughley 1987), and the availability of breeding sites (and, hence, abundance) for RRv vector species. A small scale spatial model for the sub-tropical city of Brisbane suggested that the risk of RRv disease was associated with proximity to certain vegetation types conducive to mosquito breeding, especially wetlands (Muhar et al. 2000).

Plant composition and activity in a region reflects, and can modify, local temperature, precipitation and humidity. Meteorological satellite sensors can measure these climate and vegetation variables directly (Hay et al. 1996). The normalised difference vegetation index (NDVI) is a satellite measure of vegetation "greenness", and can act as a surrogate measure for the availability of mosquito breeding sites (Garner and Kalunda 1999). The NDVI has proved a useful predictor of the incidence of several mosquito-borne diseases. Hay and others (1996) found a linear relationship between the NDVI and saturation deficit, and speculated that the NDVI may be useful in studies of arthropod vectors where saturation deficit is important. Patz and others (1998) regressed monthly NDVI with averaged weekly cases of Rift Valley fever in Kenya, and found that a combination of soil moisture modelling and satellite NDVI nearly equally predicted human biting rate.

There are some problems with using the NDVI. Monthly values have been found to be more robust than weekly indices (Patz et al. 1998). Depending on the purpose of the study, this may not capture sufficient variability in case rates. Also, cloud cover during critical (i.e., rainy) periods means that data are unavailable. Although it would have been useful to include NDVI data in this study, the data were not able to be procured at the start of this thesis. Instead, other data on vegetation were used (see Chapter 3).

1.5.6.2 The complex role of irrigation

The major natural water resources in Australia are in the tropical north and Tasmania, whereas most of the agriculture is in south-eastern Australia. In that region, the amount of water drawn from the Murray Basin for irrigation is nearly at the limit of the water resource, and the resulting environmental impacts (soil loss, salinity, acidification, and water pollution through pesticide and herbicide usage) are causing a major re-assessment of land use practices (State of the Environment Advisory Council 1996b). Irrigation methods currently used in Australia include (i) intermittent irrigation (dripper systems and micro-sprinklers) for horticulture (ii) spray or furrow for broad acre, and (iii) flood
irrigation for rice. Of these, rice irrigation, where water stays on the ground for up to six months of the year, can prolong the time in which mosquitoes breed (Gratz 1999).

The relationship between mosquito-borne disease and irrigation is complex and extremely localised. It varies with species type and composition, and the proximity of irrigated areas to human settlements. Irrigation enables mosquito breeding to occur in years and seasons when rainfall is absent or low (especially important in arid and temperate zones), and hence generally raises the probability of arboviral infection (Russell 1994). Rain falls predominantly in the winter months in most of southern-eastern Australia. Irrigation throughout the warmer summer months can provide an opportunity for breeding that would not normally exist. In dry regions, irrigation can act to elevate relative humidity and thus aid survival of vectors (Ijumba and Lindsay 2001).

Alterations to the environment can influence the dominance of vector species in unpredictable ways, however, and there is not always a positive association between irrigated regions and mosquito-borne disease transmission. A change in the available season for breeding can result in the displacement of one species with another more suited to the changed ecological conditions (Dhileepan 1996, Singh et al. 1999). Dhileepan (1996) speculated that a reduction of Aedes species abundance in one RRv enzootic region was partly attributable to modified (i.e., reduced) irrigation practices. If the new arrivals have lower vectorial capacity, or are not anthropophilic, this can even lead to a reduction in cases of disease (Ijumba and Lindsay 2001).

Although the density of mosquitoes may be higher around irrigated areas, it is the infectivity of mosquitoes (or the “transmission intensity”) in surrounding habited regions that will determine transmission rates (Ijumba and Lindsay 2001). This must in turn be related to the presence or absence of sufficient vertebrate hosts in the region (Gratz 1999). The dispersal distances of the vector species, and the extent to which recreational areas are proximal to irrigated lands, will also influence the impact of irrigation on the human infection rate.

1.5.6.3 Mosquito control

The current activities undertaken to reduce human-mosquito-virus contact in Australia are discussed in Chapter 8. What is of interest in this sub-section is the influence that mosquito control activities have on the arboviral cycle, and hence on infection of humans.
Mosquito control, when effective, has the potential to vary the magnitude of the association between climate and environment, and cases of disease. If mosquito numbers are substantially reduced, an outbreak of RRv disease may not occur even though weather conditions are suitable.

Very little information exists regarding the effectiveness of mosquito control activities in Australia. For successful vector control, the following are required: (i) incrimination of the vector species, (ii) knowledge and understanding of vector biology, (iii) surveillance (to document vector density and location, and to monitor control effectiveness), and (iv) implementation of effective control measures (Mitchell 1996). Intuitively, the effectiveness of any prevention campaign is difficult to demonstrate. Perfect prevention occurs when cases of diseases are minimal or low—and yet the absence of cases may simply signify that the disease-causing agent was absent from the environment, rather than that control measures were effective. Conversely, a large number of cases does not necessarily mean that control activities have been completely ineffectual. It may indicate that there was a successful reduction of, say, half the cases.

Ryan and others noted the poor correlation between *Ae. vigilax* densities and RRv disease cases in Maroochy Shire, and suggested "it is conceivable that selective mosquito control during epidemic seasons may depress *Ae. vigilax* light trap indices" (Ryan et al. 1999). They conclude by recommending that Local Government mosquito control programs in south-east Queensland should substantially readjust the targeting of RRv vector species to include fresh water larval habitats and additional salt water breeding species. These comments illustrate two things. First, that even in a densely populated region such as south-east Queensland, which is (comparatively) well-funded to conduct mosquito control, there is still much to be learnt about RRv vector ecology and the effective targeting of major species. Second, that confidence in the ability of selective (as opposed to blanket) control activities to significantly depress mosquito numbers in epidemic situations is, at best, weak.

Whelan and others (1996) conducted an analysis of RRv disease and environmental factors in the Darwin city region from 1990/91 to 1995/96. They regressed cases of RRv disease against rainfall, tide height, vector species density, and the number of hectares sprayed for mosquito control. Several points regarding vector control efficacy are worth making. First, the highest incidence of RRv disease was in the first year of the study when
there was “significant, although less, control effort compared with the latter years” (Whelan et al. 1996). The relatively high incidence of cases in one region, Rapid Creek, was “unexpected...as the few small sources of [vector species] were thought to be relatively well controlled by ground larval control operations”, and may suggest activity from an additional vector, *Ae. notoscriptus*, that has peri-domestic breeding habits (Whelan et al. 1996). Second, high tides would normally be expected to increase *Ae. vigilax* density. However, the relationship between high tides and RRv disease incidence, although non-significant, was negative. This suggests that some vector control activities may have been effective. Spraying was timed according to high tidal predictions. The authors concluded that the spraying of salt marshes following high tides from September to November can have some success in reducing large numbers of *Ae. vigilax* adults, whereas, a delay in spraying until December and January may destroy larvae but not sufficiently reduce the numbers of adults. Third, factors such as housing and the lifestyle of residents in different study regions were also thought to have confounded the vector control association. For example, the incidence of RRv disease was highest in one region where vector control was not practised (Litchfield), but this was also a region where the exposure of residents (from housing design and relaxation activities) was expected to have been higher.

These two studies highlight the complexity of effective mosquito control activities, in particular (i) the critical importance of the timing of spraying, (ii) the necessity for detailed and updated information about local vector type and breeding habits, and (iii) the difficulty in assessing “effectiveness”. Further important points to note are that the two study regions are population dense, both have the benefit of public health personnel with substantial training and field experience in mosquito biology and vector control activities, and the vector control measures were “targeted” rather than “blanket (i.e., they were aimed at specific species). In general then, it appears reasonable to assume that the modifying effect of vector control on the relationship between climate exposures and RRv disease in most years of Australia will be minimal. It is highly unlikely, except in very particular circumstances, to change the direction of the association. In terms of modelling, specific information about control activities would be relevant to an interpretation of results.
1.5.7 The influence of climate change on mosquito-borne diseases

1.5.7.1 Global changes to climate

Beyond the seasonal and inter-annual effects discussed so far, our changing climate - longer term natural and human-induced cycles - is likely to influence the extrinsic and intrinsic factors that drive mosquito-borne diseases. It is now widely acknowledged in the scientific community that Earth’s climate system has substantially changed since the pre-industrial era, and most of the global warming over the past 50 years can be attributed to human activities (principally the burning of fossil fuels) (IPCC 2001b). Changes have already been observed to many physical and biological systems, and there are preliminary indications that social and economic systems have also been affected.

The Intergovernmental Panel on Climate Change (IPCC 2001b) produced forty future world scenarios based on a range of possible greenhouse gas and sulphate aerosol emissions (IPCC 2000). The Special Report on Emissions Scenarios (SRES) is based on a range of assumptions about population, energy sources and regional or global approaches to development and socio-economic arrangements. The scenarios do not include any specific greenhouse gas mitigation activities. From the SRES, global temperatures are projected to increase 1.4°C to 5.8°C by 2100, relative to 1990 baseline climate. This range includes the uncertainty about how the climate system will respond to increased greenhouse gases, as well as the uncertainty about the total amount of emissions. The projected rate of warming is 0.1 to 0.5°C per decade. Sea-level is projected to rise by 9 to 88 cm by 2100, or 0.8 to 8.0 cm per decade. The observed rise over the 20th century has been 1 to 2 cm per decade. Even if global concentrations of greenhouse gases were stabilised, Earth would continue to respond to warming for many centuries (IPCC 2001b).

The increasing levels of greenhouse gas concentrations in the atmosphere are predicted to result in changes in daily, seasonal, inter-annual and decadal variability (IPCC 2001b). There are still severe limitations on the ability of global climate models to represent the complexity of observed climate variability.

Understanding how ENSO may change with global climate change is also essential for anticipating the impact of future climate on mosquito-borne diseases. Currently, climate models have mixed success in estimating inter-annual influence (Hennessy et al. 1999).
Many models predict an increase in El Niño-like conditions in the tropical Pacific (with flow-on effects for Australia), although not all do (IPCC 2001b).

**Impact of climate change on mosquito-borne disease distribution**

Projected climate change is expected to have beneficial and adverse effects on environmental and socio-economic systems, with both direct and indirect effects on human health (McMichael et al. 1996). Although assessment of the effects of climate change and its impacts still involves uncertainty, most of the effects are likely to be negative (IPCC 2001b). Potential health effects include those due to higher temperatures (e.g. increases in heat-related mortality), and to changes in temperature, humidity and rainfall patterns (e.g. changes in the incidence of inflammatory and respiratory diseases such as asthma) (IPCC 2001a).

In theory, changes in the geographical range, seasonality, and incidence rate of mosquito-borne diseases can also be expected from anthropogenic climate change, although as yet there is a lack of strong evidence that this impact has occurred (Kovats et al. 2001). It is beyond the scope of this thesis to discuss the projected future impact of climate change on the global burden of mosquito-borne diseases. The chapter on human health from the Third Assessment Report of the IPCC (2001a) provides a comprehensive scientific review of projected or observed changes in the transmission of mosquito-borne diseases. To summarise that work, temperature increases are predicted to expand the geographic range of malaria to higher altitudes, and to extend the transmission season in some locations. A possible, but less likely scenario, is that malaria transmission will also expand to higher latitudes. Such predictions assume no change in prevention effectiveness, and are for regions with limited public health infrastructure. Dengue and some mosquito-borne encephalitides may change the margins of their current distribution because of increasing climatic variability. Overall, there is likely to be a small increase in the percentage of people in the world living in regions suitable for the transmission of malaria and dengue.

**1.5.7.2 Climate change in the Australian region**

A review of the scientific evidence collected so far indicates that: (i) Australia's continental-average temperature has risen by about 0.7°C from 1910–1999, with most of this increase occurring since 1950, and (ii) minima have generally increased more than
maxima (CSIRO 2001). While Australian rainfall has varied substantially over time and space, there has been no significant continental-average trend since 1910.

Estimates of future Australian climate change have been prepared by the CSIRO (2001). Climate projections for 2030 indicate that annual average temperatures may rise by 0.4 to 2°C over most of Australia, with slightly less warming in some coastal areas and Tasmania, and the potential for greater warming in the north-west. By 2070 they may rise by 1.0 to 6.0°C. The range of warming is projected to be greatest in spring and least in winter. In the north-west, the greatest potential warming occurs in summer. Modelling results for Australia indicate that future changes in variability will be relatively small, with increases in average maximum and minimum temperature mainly determining the change in extremes.

Most models simulate an increase in extreme daily rainfall leading to more frequent heavy rainfall events and flooding. This can occur even in regions where average rainfall decreases. Reductions in extreme rainfall are projected where average rainfall declines significantly. Projected annual averages tend towards a decrease in the south-west (-20% to +5% by 2030 and -60% to +10% by 2070), and in parts of the south-east and Queensland (-10% to +5% by 2030 and -35% to +10% by 2070). In some other areas, including much of eastern Australia, projected ranges are -10% to +10% by 2030, and -35% to +35% by 2070. These extremes hide a seasonal re-distribution of rainfall, where winters and springs are projected to become drier, and summers wetter. The ranges for the tropical north (-5% to +5% by 2030 and -10% to +10% by 2070) represent little change from current conditions.

Impact of climate change on mosquito-borne disease distribution

In 2003, the Australian Government published a report titled Human health and climate change in Oceania: a risk assessment (McMichael et al. 2003). That assessment involved a review of the current state of knowledge regarding the impact of projected climate change on mosquito-borne disease distribution and prevalence in the Australian region, including prior work undertaken for malaria (Bryan et al. 1996, Walker 1998b), dengue (Mackenzie et al. 1996a), and for arboviruses in general (Russell 1998b). Models of future distribution of mosquito-borne diseases in the region under different climate change scenarios had been prepared for dengue (de Wet et al. 2001, Hales et al. 2002b) and malaria (Bryan et al. 1996).
In addition to the review, the authors conducted a quantitative analysis of the potential health impacts of climate change in the medium-term (2030 and 2070) in Australia and, in specified instances, the neighbouring populations of New Zealand and the Pacific Islands. In relation to mosquito-borne diseases, the key findings from the risk assessment were:

- The “malaria receptive zone” may expand southwards down the coast of Queensland, to include regional towns such as Rockhampton, Gladstone and Bundaberg. However, in the foreseeable future malaria itself is not a direct threat to Australia under climate change, as long as a high priority is placed on prevention via the maintenance and extension of public health and Local Government infrastructure.

- Suitable conditions for the transmission of dengue may expand southwest down to Carnarvon, and southeast down to Maryborough and Gympie by 2050. If no other contributing factors were to change, a larger number of people living in northern parts of Australia would be at risk of dengue infection (a total of 0.3-0.5 million in 2020, and 0.8-1.6 million in 2050). This increased risk need not mean an increase in dengue cases, provided there is (i) continuing expansion of vector control and public health surveillance, and (ii) quarantine efforts to ensure that a secondary dengue vector, *Ae. albopictus*, does not become established in the country.

- In relation to RRv disease, rising temperatures combined with changing rainfall patterns are likely to have significant impacts on transmission that would vary by geographical area. Russell has speculated that *Aedes* populations in dry areas may be adversely affected by decreased winter rainfall, possibly delaying or precluding virus activity (Russell 1998b). Conversely, the predicted increase in summer rainfall may increase the availability of mosquito habitat that, combined with higher average temperatures, may lead to higher humidity, a lengthened season of abundance and greater transmission levels (Russell 1998b). Rising temperatures on their own, without an accompanying increase in rainfall in a region, are unlikely to lead to an increase of RRv disease in most parts of Australia. In view of the complex dynamics of transmission for this disease, more research is required at the regional level into the ecology of the virus, its hosts and vectors, and the impact of human activities (such as environmental modification, vector control, and preventive education) on reducing the risk of infection.

In summary, climate change is expected to lead to changes in the pattern and incidence of mosquito-borne disease in Australia, as in the rest of the world. Some mosquito-borne
Figure 1.4  Representation of the determinants of Ross River virus disease.

**Climate and weather**
(rainfall, temperature, humidity, tidal inundation)

**Vertebrate host**
(breeding and feeding)

**Mosquito vector**
(breeding, survival, abundance, longevity, dispersal)

Factors moderating amplification:
- vertebrate immunity levels
- vector species type

**Transmission via mosquito bite**

Factors moderating transmission:
- human immunity levels (in-migration of non-immunes, tourists)
- human proximity to mosquito habitat (siting of residential buildings)
- housing design (screens, air conditioning)
- human behaviour (lifestyle, personal protection)
- vulnerability (economic disadvantage, health-seeking behaviour)
- public health focus on disease prevention (early warning mechanism: vector surveillance, vector control, prevention education, disease monitoring)

**Virus amplification**

**RRv infection**

**RRv disease notification**

Number of cases affected by:
- sub-clinical to clinical ratio

Number of notifications affected by:
- under and over reporting
- spatial misclassification
- false positive rate

Other determinants of mosquito breeding and survival:
- irrigation, constructed wetlands
- effective vector control
- insecticide resistance
diseases with relatively simple transmission cycles (such as malaria and dengue) have been amenable to study, and cautious predictions of future changes have been proposed. The relative complexity of the RRv transmission cycle, and the numerous regional ecologies of the virus, have hampered predictions thus far. A first step would seem to involve an examination of the relationship between current meteorological factors and disease outbreaks in ecologically homogeneous regions. This may help in predicting the future consequences of climate change.

1.5.8 Summary of the determinants of RRv disease

Figure 1.4 provides a graphical summary of the major determinants of RRv disease in humans.

Factors promoting outbreaks of RRv disease are:

- Climate and weather conditions conducive to (i) the breeding, abundance, survival, and longevity of amplifying and transmission vectors, (ii) the breeding and abundance of vertebrate hosts, and (iii) the replication of the virus. The most important influences are high rainfall, high temperatures and humidity, and high tides. Vegetation can provide a proxy measure of these.
- Sufficient numbers of competent amplifying and transmission vectors.
- Sufficient numbers of susceptible vertebrate hosts.
- Non-immune humans.
- Environmental factors that provide semi-permanent breeding sites (such as irrigation or other forms of water management).
- Human incursion into the natural RRv transmission cycle.

Factors that lessen or restrict RRv disease outbreaks are:

- Periods of cool weather, low rainfall, and low tides.
- A build-up of herd immunity in the vertebrate host population.
- Displacement of vectors by others that have lower transmission efficiency, or exclusive zoophilic feeding preferences.
- Good public health infrastructure (such as effective vector control, surveillance of vectors, disease reporting).
- Human behaviour reducing exposure (such as housing design, personal protection, occupation and recreation).
1.6 CONCLUSIONS

Mosquito-borne diseases represent a large burden of disease at the global level, and in Australia. The current evidence suggests that, even if public health response is optimal, the combination and scale of climate change and other global changes will lead to an increase in mosquito-borne disease incidence in the future. This changing incidence will be socially, economically, politically, and biophysically mediated.

Climate and weather are prime influences on RRv epizootic and enzootic activity. RRv has a relatively complex transmission cycle, with multiple vectors and vertebrate hosts involved. Long-term and inter-annual climate constrain the distribution of the vectors, and the seasonality of transmission. Rainfall, temperature, humidity, and tidal inundation appear to strongly regulate the transmission of the virus between vector and vertebrate host populations.

There are a number of different spatial and temporal patterns of RRv disease across the Australian continent. These appear to be overwhelmingly related to differences in the ecological conditions that drive vector and virus proliferation. Outbreaks have not been recorded in the winter period in the cooler parts of the country. In tropical regions, where temperature and humidity are relatively constant, cases occur throughout the year. Patterns of disease in arid and temperate regions are epidemic, with outbreaks typically occurring after abnormally high rainfalls and in non-consecutive years. Regional differences in human behaviour, public health response, and environmental modification can at times moderate these patterns (by changing the magnitude of effect).

Notwithstanding the above conclusions, the relationship between RRv disease outbreaks and climate requires further exploration. Extensive reports and studies indicate the coincidental occurrence of weather events and disease, but the temporal and spatial relationships have not been quantified. A number of epidemiological studies have considered single variable associations with RRv disease (i.e., Harley and Weinstein 1996, Lindsay et al. 1996b, Tong et al. 1998, Tong and Hu 2001, Tong et al. 2002) or have developed predictive models based on single variables (Maelzer et al. 1999). These studies have not accounted for the interactive effect of multiple meteorological factors on disease transmission. They also have not examined the contribution of vertebrate host immunity to the occurrence of outbreaks (critical, in particular, for temperate and arid regions).
A biologically rational study of the relationship between RRv disease outbreaks and climate requires an examination of the appropriate scale and boundaries for analysis. The different ecologies of the virus between arid, temperate, and tropical regions, and between coastal and inland areas, suggests the need to conduct analyses within climatically similar regions, rather than at State or continental levels. The appropriate scale should be shaped by (i) an understanding of the main climatic drivers of viral activity, and their influence on RRv ecological boundaries, and (ii) consideration of the purpose of the research.

The value of climate as a predictive tool for RRv disease outbreaks has not been examined. An ability to predict epidemics could enable the more targeted use of prevention and control strategies. If meteorological variables are to be useful for public health response, then the timeliness of prediction, and their role in conjunction with other indicators in an early warning system, are important to determine.
Chapter 2

Research questions and thesis synopsis

2.1 RESEARCH QUESTIONS

To summarise from Chapter 1, there is substantial evidence linking mosquito abundance and viral infectivity to certain types of weather. Although climate and weather are not the only determinants of outbreaks of mosquito-borne diseases, they are primary ones. This thesis takes an epidemiological approach to bring together data on disease, demography, climate, and the environment to study the interactions between climate variability and RRv disease patterns.

Some of the questions that helped frame the direction of this thesis were:

- A few population-dense public health and local government regions in Australia are well resourced for RRv disease management. However, most have minimal to no resources for control or prevention activities. How could RRv surveillance be simply and cheaply improved?
- Which climate factors are most useful for prediction - or the cheapest and easiest to monitor - from a public health perspective?
- At what scale would analyses generate the best predictive model (i.e., in terms of model validity and public health utility)?
From these questions, the aim and objectives for this thesis were established:

The **aim** was to determine if climatic and environmental data can be used to predict epidemics of RRv disease with sufficient accuracy and timeliness to be useful for public health surveillance.

- **The primary objective** was to develop predictive models for the probability of RRv disease occurrence in two different study areas of Australia.
- A **secondary objective** was to develop a geographical classification system based on the main climatic and environmental factors that influence the ecology of RRv.
- A **third objective** was to propose a structure for an RRv disease early warning system that could be used to improve the effectiveness of public health response (particularly mosquito control and public notification).

To accomplish these objectives, I sought to use low cost, readily available data and simple methods that could be used by the majority of public health authorities.

### 2.2 Thesis Synopsis

Chapter 1 describes the global and Australian burden of mosquito-borne disease, and makes a case for focusing this thesis on the mosquito-borne disease with the major public health burden in Australia - Ross River virus disease. It reviews the scientific literature on the transmission cycle of the disease, including the ecology of the virus, its vertebrate hosts (predominantly macropods), and the major mosquito species implicated as vectors. The effect of the virus on humans is discussed, along with known risk factors. This chapter also presents a review of the evidence for the influence of weather, climate, and environment (such as vegetation, irrigation, and mosquito control) on RRv transmission.

The chapter concludes by considering the gaps in research on the relationship between RRv disease and climate. It argues that previous work has not accounted for the interactive effect of multiple meteorological factors or the role of vertebrate host immunity in analyses of RRv disease outbreaks. Nor has there been consideration of the use of prediction as a forecast tool for public health management of RRv disease.

Chapter 3 presents the disease and exposure data that were used in the study. Primary variables for analysis were climate attributes that most strongly affect mosquito proliferation.
Chapter 4 is an analysis of the epidemiology of RRv disease in Australia for an eight-year period (1991 to 1999). This chapter shows that analyses of the temporal and spatial distribution within rainfall zones is preferable to one at the State level. Even so, there is substantial variation in notification rates within the zones, and analyses at a finer resolution – the Statistical Local Area (SLA) – better describes the patterns.

Chapter 5 describes the method used to define the regions for analysis, and the criteria used to decide which areas would be studied. It presents the methodology that was used for the exposure-disease modelling in the two case study areas (the Murray River of south-eastern Australia and the south-west of Western Australia), including the management of bias and confounding.

Chapters 6 and 7 are the main experimental chapters. Chapter 6 presents the findings from the RRv disease epidemic modelling, using only climate data as explanatory variables. Two regions surrounding the Murray River in south-eastern Australia were studied - one temperate, and one semi-arid. This chapter reviewed the published literature of outbreaks and vector studies across the Murray River regions, and the evidence was used to generate hypotheses about the relationship between climate and the transmission of RRv disease. Chapter 7 describes the modelling of RRv disease epidemics in a temperate region in the south-west of Western Australia. It replicates the methods used in the Murray Valley study area, to determine if they could be successfully repeated in another area. This chapter also tested whether mosquito surveillance data could enhance the sensitivity of predictive models, in addition to the contribution of the other exposure variables (i.e., climate and environment).

Chapter 8 considers the prospects for using such models in an RRv early warning system. It reviews existing health-related early warning systems, and proposes a number of generic requirements for a system that could work to improve the risk management of mosquito-borne diseases. The models from the two study areas are assessed against these requirements, and their potential benefits and limitations are discussed.

Chapter 9 summarises the main conclusions of the thesis, and suggests areas for further research. Appendix A contains a list of an award and the papers that have resulted from this thesis, and of relevant presentations made to national and international meetings.
3.1 Introduction

This chapter describes the type and structure of the disease, demographic, climatic, and environmental data used to model Ross River virus disease (RRV disease) epidemics. It discusses the quality of these data, and limitations in their coverage.

3.2 Disease data

3.2.1 Type and structure

Notifications of RRV disease were obtained from the National Notifiable Diseases Surveillance System (NNDSS), with permission from the Communicable Diseases Network of Australia (1999). The NNDSS was established in 1990, and is responsible for coordinating the national surveillance of more than 40 communicable diseases or disease groups, of which RRV disease is one. Comprehensive national reporting of RRV disease did not start until 1991. Cases are reported to State Health Departments, entered onto a database, and passed to the Commonwealth Department of Health and Aged Care for compilation on a fortnightly basis. This study used all notifications for the period July 1991 to June 1999 (n=39 422). There were no reports from Tasmania for the first three years of the study (1991/92 to 1993/94).

The disease dataset contained anonymous information about each case in an individual record format. Table 3.1 gives the data structure. Fields were age, sex, postcode of residence, and State or Territory code of diagnosis. Also included were the day, month,
and year when symptoms were estimated to have commenced (as determined by the consulting doctor in conjunction with the patient), and the day, month, and year when the case was notified to the relevant health authority (either by the doctor or the pathology laboratory, following a confirming diagnosis). The Northern Territory Health Department records the location of all communicable diseases (including RRv disease) by their health district code, rather than by postcode (discussed further in Chapter 4).

Table 3.1 Structure of the RRv disease data set.

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Description</th>
<th>Status</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record number</td>
<td>–</td>
<td>As supplied</td>
<td>1 ... n+1</td>
</tr>
<tr>
<td>States</td>
<td>State or Territory code</td>
<td>As supplied</td>
<td>1-8</td>
</tr>
<tr>
<td>Postcode</td>
<td>–</td>
<td>As supplied</td>
<td>–</td>
</tr>
<tr>
<td>Age group</td>
<td>5 year categories</td>
<td>Derived from ‘age’</td>
<td>0-4, 5-9, ..., 95-99</td>
</tr>
<tr>
<td>Sex</td>
<td>Males=1, females=2, unknown=9</td>
<td>Re-coded from ‘sex’</td>
<td>1, 2, 9</td>
</tr>
<tr>
<td>Onset</td>
<td>Estimated start date of symptoms</td>
<td>As supplied</td>
<td>dd/mm/yy</td>
</tr>
<tr>
<td>Report</td>
<td>Date of report to Health Department</td>
<td>As supplied</td>
<td>dd/mm/yy</td>
</tr>
<tr>
<td>RRv year</td>
<td>1=01/07/1991-30/06/92, ..., 8=01/07/1998-30/06/99</td>
<td>Derived from ‘onset’</td>
<td>1-8</td>
</tr>
<tr>
<td>RRv month</td>
<td>1=Jul, 2=Aug, 3=Sep, ..., 12=Jun</td>
<td>Derived from ‘onset’</td>
<td>1-12</td>
</tr>
<tr>
<td>X coordinate</td>
<td>Decimal longitude values</td>
<td>Derived from ‘postcode’</td>
<td>–</td>
</tr>
<tr>
<td>Y coordinate</td>
<td>Decimal latitude values</td>
<td>Derived from ‘postcode’</td>
<td>–</td>
</tr>
<tr>
<td>ASGC</td>
<td>Australian Standard Geographical Classification</td>
<td>Derived from the ‘xcoord’ fields</td>
<td>–</td>
</tr>
</tbody>
</table>

I attempted to obtain RRv disease data prior to 1991. From 1982 to 1999 monthly notifications of RRv disease and other diseases were reported to the Virology and Serology Laboratory Reporting Scheme (LabVISE) – coordinated by the Commonwealth Department of Health and Aged Care. These data were supplied on a voluntary basis by self-selected laboratories from all States. For the 18-year period, 22 167 RRv disease notifications were recorded. Unfortunately, the LabVISE data were subject to several problems:

- The number of laboratories participating in the scheme varied over time, resulting in a lack of continuity.
Chapter 3 - Data collection

- The location of the participating laboratories was arbitrary. Some State and Territory jurisdictions were better represented than others (Communicable Disease Surveillance 1997).
- There were no geographical references for the cases (such as postcode or State of residence). Although there was a “State” field in the data set, it referred to the location of the laboratory not the case. It is often the case that blood specimens are collected in one State and then freighted to laboratories in other States, depending on commercial affiliations or “catchment” areas. Hence, even the “State” field could not reliably be used to provide a measure of State-wide trends in relation to seasonality of infection, or the identification of “high” or “low” RRv disease years.
- The records did not include any individual information such as age or sex.

In light of these defects, the LabVISE data were not used in analysis.

Reports of RRv disease outbreaks were sourced from the peer-reviewed and “grey” literature. The methods used were:
- cross-referring from more recent publications to older ones not listed on databases,
- a search of Medline and Commonwealth Scientific and Industrial Research Organisation (CSIRO) libraries,
- a search of annual reports and newsletters from State and Territory health departments, and
- a search of newsletters and conference proceedings of the Mosquito Control Association of Australia and Arbovirus Research in Australia.

This information was used to build a profile of the types of weather conditions preceding the major outbreaks.

I also contacted all State and Territory Health Departments by letter, and followed up with a phone call, to obtain information on the status of RRv disease case follow-up data in each State, and on risk management approaches.

3.2.2 Disease data quality

3.2.2.1 Inconsistencies with the case definition

The diagnostic methods for determining arboviral infection include haemaglutination inhibition, neutralisation, complement fixation, and enzyme-linked immunosorbent assay (ELISA). ELISA testing allows measurement of immunoglobulin M (IgM) and IgG
separately (Mackenzie et al. 1993). IgM antibody levels can be detected in blood sera within 48 hours by ELISA. IgM provides a “presumptive diagnosis” of recent infection, because it is produced early after infection and usually does not persist at high titre (Carter et al. 1985). However, IgM antibody can persist for months or years after infection in some people (Carter et al. 1985, Smith et al. 1993), and persistence at 18 months is not uncommon (Linda Hueston pers. comm. 2000). Many people experience delayed symptoms of RRv disease, or show no symptoms at all. The presence of IgM may, in some cases, represent past infection or another illness (Flexman et al. 1998). The possibility of misdiagnosing a past infection as a recent one increases when IgM detection is the only test used (Vitarana et al. 1994).

Various laboratories throughout Australia undertake arboviral testing and subsequent reporting of human infections (both commercial and government-funded), but uniformity has not been reached between laboratories regarding a standardised testing protocol (Mackenzie et al. 1993, Linda Hueston pers. comm. 2000). The Communicable Diseases Intelligence Laboratory Reporting Scheme has attempted to standardise case definitions for routinely notified data. The Scheme defined a confirmatory diagnosis of RRv disease to be (i) a four-fold rise or drop in antibody titre between the acute- and convalescent-phase sera, and (ii) where the two serum specimens are obtained 10-14 days apart (Mackenzie et al. 1993). In addition, acute phase serum should be collected within 7 days, and convalescent-phase serum within 8 to 28 days, of illness onset. A presumptive case is one for which only acute-phase serum (with IgM antibody) is available (Mackenzie et al. 1993).

Regardless, it is routine for many laboratories to make a positive diagnosis of RRv infection on the basis of a single titre (Linda Hueston pers. comm. 2000), and it is recognised that the majority of RRv notifications in Australia are presumptive (Curran 1994, Curran 1996). Some logistical problems work to prevent the application of the “two specimen diagnosis” protocol, the main of which is the fact that many patients present to their doctor after acute conditions resolve (from the analysis in Chapter 4, this may be in the order of 20-40%). For Barmah Forest virus disease, which has a similar clinical picture, van Buynder and others (1995) have observed that few clinical practitioners collect a second specimen in situations where there is both a positive IgM confirmation on the first specimen and a consistent clinical picture for the disease.
For the majority of cases, a presumptive diagnosis may be acceptable provided the IgM is detected by a method of proven specificity (Rich et al. 1993). Until recently, however, most testing of RRv infection in Australia was done with the commercial PanBio indirect ELISA kit (a second kit manufactured by Biocene is now also available: Harley et al. 2001). The PanBio kit has been reported to have low specificity in several studies. Rich and others (1993) reported 14% of diagnosed cases were false positives, and Lloyd and others (2001b) found 20% were false positives in their study. The small potential for cross-reaction between RRv and other flaviviruses also complicates specificity. There is a close association between BFv disease and RRv disease symptoms (Boughton et al. 1988, Phillips et al. 1990, van Buyneder et al. 1995, Flexman et al. 1998). According to Rich and others (1993), cross-reaction is “frequent” with some ELISA tests (no estimation of this is provided). BFv disease notifications accounted for 8% of all arboviruses that were notified during the study period, while RRv disease accounted for 85% (Communicable Diseases Network Australia 2002). It is possible that some presumptive cases of RRv disease were misdiagnosed as BFv disease (or vice versa), although there was no way of determining what this amount may have been.

### 3.2.2.2 Reporting quality

Due to under-reporting, notified cases mostly represent only a proportion of the total number of cases that occurred (Curran et al. 1997). This proportion will vary between States and Territories, and over time. Methods of surveillance vary between jurisdictions, each with different requirements for notification by medical practitioners, laboratories and hospitals (Curran et al. 1997). Notification data are likely to be biased towards cases with typical clinical symptoms (Tai et al. 1993). As well, people with less severe illness, including children, may not seek medical help, or may be mis-diagnosed. Under-reporting of RRv disease has been noted in regions where the disease is endemic, and during outbreaks when, paradoxically, a decrease in laboratory notifications has been reported (Russell 1998a). This may be due to the behaviour both of patients (who become informed of symptoms as an outbreak develops, and choose not to attend a practitioner due to the lack of specific treatments or a cure for the disease) and practitioners (who, confronted with multiple cases of apparent RRv disease, diagnose from clinical symptoms rather than send blood samples for serological confirmation). Under-reporting is unlikely to impact substantially on this study. The outcome variable for analysis was an epidemic of RRv disease, rather than case incidence (rationale discussed in Chapter 5).
Epidemics are likely to be represented through the NNDSS data with reasonable accuracy.

A potential source of bias was the difference between place of infection and place of residence. Cases were notified by their postcode of residence only (i.e., not the presumed place of infection). Several authors have observed that the geographical reporting of RRv disease fairly accurately reflects the place where the disease was acquired (Hawkes et al. 1985, Selden and Cameron 1996). However, there are some situations where this could not be true. In particular, RRv infections are likely to be under-estimated in RRv active coastal areas (in which regions holidaying city residents become infected, but do not report to a physician until returning home). Conversely, over-estimation of RRv disease is common in metropolitan areas, due to imported cases that have been infected elsewhere (Amin et al. 1998). Statistical Local Areas (SLAs) in rural parts of the country are very much larger than in cities. To minimise this bias, I assumed that work and recreation (and hence infection) for residents of rural SLAs would generally occur within the same SLA. This could not be assumed for metropolitan SLAs, and consequently these areas were not selected for the predictive modelling.

Case follow-up provides an opportunity to estimate the place of infection. Typically, confirmed cases are contacted either directly by Health Department officials, or are indirectly invited to complete a form sent by the laboratory with the positive diagnosis. These data are used in a few States to determine “hot spots” of disease and to guide mosquito control activities. New South Wales and Victoria generally follow-up cases with a metropolitan postcode to establish any travel history to RRv active areas, but do not generally attempt to follow-up cases living in RRv active areas. In the Murray River region, cases are sometimes followed-up at the commencement of the RRv season. Only two States – South Australia and Western Australia – attempt to conduct routine follow-up of RRv disease cases. The contact rate varies, depending on the size of the outbreak and the time of the season. At the beginning of the season there is more incentive to identify the place of infection, as the information can be used to guide vector control activities. Follow-up during very large outbreaks can tax the resources of public health units, and not all cases will be contacted as diligently in these circumstances. Dr Michael Lindsay from the Health Department of Western Australia has collected RRv case follow-up data for the period 1991/92 to 1999/2000, for three mosquito-trapping regions in the Southwest study area (approximately nine of the fourteen SLAs in the study area). Unfortunately, the low contact rate (less than 50% of the notifications reported to the
NNDSS in some years and SLAs), and the fact that cases across the whole study area were not followed-up, meant these data were not used for this study.

### 3.3 Demographic Data

Demographic information for the period was obtained from the Australian Bureau of Statistics' Integrated Regional Database, version 98 (Australian Bureau of Statistics 1998a). I obtained estimated annual resident population by sex and five year age category by SLA (n=1272) for 1996, based on the Australian Standard Geographical Classification Version 2.6. A fuller description of the SLA classification system is provided in Chapter 4.

### 3.4 Climate and Environment Data

The following section details the exposure data used in the analysis. The climate and environment factors listed in Table 3.2 have been identified as major determinants of the life-cycle of RRv, its vectors, and vertebrate hosts (see Chapter 1). For this thesis, it is assumed that the entire populations of RRv vectors and vertebrates in an SLA are equally exposed to these factors (unless otherwise noted).

<table>
<thead>
<tr>
<th>Virus, vector, and vertebrate dynamics</th>
<th>Climate and environment factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breeding</td>
<td>- Rainfall (timing, intensity, duration)</td>
</tr>
<tr>
<td></td>
<td>- Climatic indices known to affect rainfall (SOI, SST)</td>
</tr>
<tr>
<td></td>
<td>- Tidal inundation (absolute and mean tide height)</td>
</tr>
<tr>
<td>Abundance</td>
<td>- Rainfall (timing, intensity, duration)</td>
</tr>
<tr>
<td></td>
<td>- Climatic indices known to affect rainfall (SOI, SST)</td>
</tr>
<tr>
<td></td>
<td>- Temperature (maximum, minimum, absolute maximum, absolute minimum, average)</td>
</tr>
<tr>
<td>Spread</td>
<td>- Wind speed</td>
</tr>
<tr>
<td>Survival</td>
<td>- Temperature (extreme cold reduces larval reproduction; extreme heat causes desiccation)</td>
</tr>
<tr>
<td></td>
<td>- High humidity (increases survival)</td>
</tr>
<tr>
<td>Viral replication</td>
<td>- High temperatures (reduce extrinsic incubation period)</td>
</tr>
</tbody>
</table>
3.4.1 Climatic indices

The major inter-annual cycle that affects Australian weather patterns is the El Niño-Southern Oscillation (ENSO). Two indices of ENSO were used as variables in analysis: the Southern Oscillation Index (SOI) and sea surface temperature. The Climate Prediction Center (2001) classification for the strength of La Niña and El Niño episodes was used for the study period (Table 3.3).

3.4.1.1 Southern Oscillation Index

The SOI is a measure of the standardised air pressure difference between Tahiti and Darwin, eastern and western points of the Pacific Basin. El Niño events are associated with strong negative values of the index, and La Niña events with strong positive values. The relationship between an El Niño or La Niña year and the SOI is not perfectly correlated, although it is very strong (Kovats et al. 1999).

<table>
<thead>
<tr>
<th>Year</th>
<th>JFM</th>
<th>AMJ</th>
<th>JAS</th>
<th>OND</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>N</td>
<td>N</td>
<td>W-</td>
<td>W-</td>
</tr>
<tr>
<td>1991</td>
<td>W-</td>
<td>W-</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>1992</td>
<td>W+</td>
<td>W+</td>
<td>W-</td>
<td>W-</td>
</tr>
<tr>
<td>1993</td>
<td>W-</td>
<td>W</td>
<td>W</td>
<td>W-</td>
</tr>
<tr>
<td>1994</td>
<td>N</td>
<td>N</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>1995</td>
<td>W</td>
<td>N</td>
<td>N</td>
<td>C-</td>
</tr>
<tr>
<td>1996</td>
<td>C-</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>1997</td>
<td>N</td>
<td>W</td>
<td>W+</td>
<td>W+</td>
</tr>
<tr>
<td>1998</td>
<td>W+</td>
<td>W</td>
<td>C-</td>
<td>C</td>
</tr>
<tr>
<td>1999</td>
<td>C+</td>
<td>C</td>
<td>C-</td>
<td>C</td>
</tr>
</tbody>
</table>

Source: (Climate Prediction Center 2001)

SOI data were sourced from CSIRO Division of Atmospheric Research (1999). Indices that reflect the variability of the Southern Oscillation were available from several sites. The data from the Division of Atmospheric Research were used because of work undertaken by that organisation to improve the interpolation of missing historical values in the data set (Janette Lindesay pers. comm. 2000).
I obtained monthly SOI for the study period and for one year prior (to allow for lagged calculations). The formula used to calculate the monthly SOI was:

\[
\text{Southern Oscillation Index} = 10 \times \left[ \frac{(T-D) - (T-D)_{\text{ave}}}{\text{sd}(T-D)} \right]
\]

where,

\[
T-D = \text{monthly Tahiti pressure minus monthly Darwin pressure}
\]

\[
(T-D)_{\text{ave}} = \text{long-term monthly average of T-D (from January 1933 to December 1992)}
\]

\[
\text{sd}(T-D) = \text{long-term monthly standard deviation of T-D (from January 1933 to December 1992)}.
\]

3.4.1.2 Sea surface temperature

Seasonal sea surface temperature is another factor that influences Australian rainfall, independent of the Southern Oscillation (Nicholls 1989). Sea surface temperature anomalies are defined as deviations for a specified region from the average for the years 1961-1990, a standard recommended by the World Meteorological Organization (Kovats et al. 1999). The Niño 3.4 is an area in the equatorial Pacific (170°-120°W and 5°N-5°S) where the “Niño 3” and “Niño 4” regions meet, and where the sea surface temperatures peak most during an El Niño event. This region is the current standard used for measuring changes in sea surface temperature. I obtained monthly data from the Climate Prediction Center (1999), part of the United States National Oceanic and Atmospheric Association. I calculated anomalies for this index as the difference between a monthly value and the long-term average value for that month (based on the period 1950 to 1979).

3.4.2 Weather data

Weather data for the period were available in two formats: as “point observations”, or as an “interpolated surface”. Point observations are collected by the Australian Bureau of Meteorology from weather stations across the country. They comprise the daily rainfall record, and the daily record of temperature, humidity, and other phenomena. An interpolated surface comprises point observations from which the digital weather “surface” is derived. This forms a grid of the country, with data values in each grid cell.

Observational datasets are usually spatially and temporally incomplete. There are significant problems with constructing a continuous, reliable weather data record from point observations for a project of this type, summarised by Jeffrey and others (2001) as follows:
Data may be recorded for discrete periods, not spanning the time period of interest. There may be intermittent gaps in data collection during the period. The data may contain systematic or random errors. In addition, weather stations are arbitrarily placed in terms of the needs of this study, and are not systematic as regards geographic features or distance between stations.

The Queensland Department of Natural Resources (2000) has derived an interpolated surface for the whole of Australia, based on observational data collected by the Bureau of Meteorology from the 1880’s onwards. I used the interpolated surface data for this study because it provided national coverage, and the data were of comparatively high quality (discussed below). Another advantage of the Department of Natural Resources data set was that daily updates can be downloaded free from the World Wide Web (http://www.bom.gov.au/silo/). This fact satisfied an objective of the thesis – to choose data and methods that could be accessed and used by the average public health area unit.

3.4.2.1 Interpolated weather surface data

The spatial weather data were purchased under a research licence agreement from the Queensland Department of Natural Resources. Data were obtained for the whole of the country for July 1990 to June 1999 inclusive (the extra year’s data at the beginning of the study allowed for analysis of lag relationships). The data were at 0.05 degree (i.e., approximately 5 km²) grid cell resolution. The surface was an estimate of the expected weather value in each grid cell for each time point, based on a weighted algorithm of surrounding recorded values and the long-term historical record.

The temporal unit of observation for the study was a month. Weather data - either average (e.g. an average of daily temperatures) or count values (e.g. number of rain days) – were summarised into monthly climate variables. The spatial unit of observation was an SLA. A summary value was calculated for each SLA, for each month period. For example, to derive the climate variable “mean maximum temperature”, (i) the maximum temperature values per day in a month were interpolated to each grid cell within an SLA, (ii) an SLA mean was calculated for each separate day of the month, and (iii) an overall monthly figure was obtained from the mean of the 31 (30, 29, or 28) daily SLA values.
The climate variables that were used in analysis, their units of measurement, and methods for calculating the spatial averages are outlined below.

Rainfall

In Chapter 1, three aspects of rainfall were identified as critical drivers of arbovirus disease outbreaks. These were the timing of an event (i.e., month, season), the total amount of rainfall, and the duration of rainfall. Accordingly, data were obtained for the total amount of rainfall, and the number of rain days in a month. The number of rain days provides a proxy measure for the duration of rainfall.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Method for calculating monthly SLA value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total rainfall</td>
<td>mm</td>
<td>SLA average of (sum of daily rainfall per pixel)</td>
</tr>
<tr>
<td>Number of rain days</td>
<td>count</td>
<td>Sum of number of rain days in an SLA. A “rain day” occurred when &gt;=20% of pixels recorded &gt;=1mm rain.</td>
</tr>
</tbody>
</table>

Temperature

Most temperature observing stations (in the year 2000 there were ~700 across Australia) record two readings per day. Some record three hourly readings, and a very few (those near metropolitan centres) record hourly readings.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Method for calculating monthly SLA value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean temperature</td>
<td>°C</td>
<td>SLA average of (sum of daily mean temperature values per pixel / number of days in the month). Daily mean temperature = (maximum temperature + minimum temperature)/2</td>
</tr>
<tr>
<td>Mean maximum temperature</td>
<td>°C</td>
<td>SLA average of (sum of daily maximum temperature values per pixel / number of days in the month)</td>
</tr>
<tr>
<td>Absolute maximum temperature</td>
<td>°C</td>
<td>SLA average of (highest daily maximum temperature value per pixel in the month)</td>
</tr>
<tr>
<td>Mean minimum temperature</td>
<td>°C</td>
<td>SLA average of (sum of daily minimum temperature values per pixel / number of days in the month)</td>
</tr>
<tr>
<td>Absolute minimum temperature</td>
<td>°C</td>
<td>SLA average of (lowest daily minimum temperature value per pixel in the month)</td>
</tr>
</tbody>
</table>
Evaporation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Method for calculating monthly SLA value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaporation</td>
<td>mm</td>
<td>SLA average of (sum of daily evaporation values per pixel / number of days in the month)</td>
</tr>
<tr>
<td>Maximum evaporation</td>
<td>mm</td>
<td>SLA average of (sum of daily maximum evaporation values per pixel / number of days in the month)</td>
</tr>
<tr>
<td>Minimum evaporation</td>
<td>mm</td>
<td>SLA average of (sum of daily minimum evaporation values per pixel / number of days in the month)</td>
</tr>
</tbody>
</table>

Humidity

Relative humidity is the ratio of the actual vapour pressure to the saturation vapour pressure, expressed as a percentage. The amount of water vapour the air can hold increases with temperature. Relative humidity therefore decreases with increasing temperature if the actual amount of water vapour stays the same. I tested the predictive benefits of three different temperature points: maximum, minimum, and average.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Method for calculating monthly SLA value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative humidity at average temperature</td>
<td>%</td>
<td>SLA average of (sum of daily average relative humidity values per pixel / number of days in the month)</td>
</tr>
<tr>
<td>Relative humidity at maximum temperature</td>
<td>%</td>
<td>SLA average of (sum of daily maximum relative humidity values per pixel / number of days in the month)</td>
</tr>
<tr>
<td>Relative humidity at minimum temperature</td>
<td>%</td>
<td>SLA average of (sum of daily minimum relative humidity values per pixel / number of days in the month)</td>
</tr>
</tbody>
</table>

Vapour Pressure

Vapour pressure is a measure of the sub-pressure water vapour in the atmosphere, derived from wet/dry bulb readings.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>Method for calculating monthly SLA value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average vapour pressure</td>
<td>hPa</td>
<td>SLA average of (sum of daily average vapour pressure values per pixel / number of days in the month)</td>
</tr>
<tr>
<td>Maximum vapour pressure</td>
<td>hPa</td>
<td>SLA average of (highest of the daily vapour pressure values in a month per pixel)</td>
</tr>
<tr>
<td>Minimum vapour pressure</td>
<td>hPa</td>
<td>SLA average of (lowest of the daily vapour pressure values in a month per pixel)</td>
</tr>
</tbody>
</table>
3.4.2.2 Quality of the interpolated surface

A comprehensive overview of the spatial interpolation method used to construct the Queensland Department of Natural Resources data set has been compiled by Jeffrey and others (2001). In summary, monthly rainfall totals were interpolated using Hutchinson's spline (Hutchinson 1995). This technique is best when applied to long-term mean surface, but appears to be reasonable for time steps of as short as one month (Jeffrey et al. 2001). In addition to a straight interpolation between observed values, data were transformed by the long-term mean and variance of monthly rainfall at each point, at each time of year, and so interpolation was performed on perturbations from a mean. This improved the interpolation by including information contained in the long-term surface, which was built from a much greater number of stations than would have been available in any one month (such as information relating to topographic influences on rainfall).

For the other climate variables (temperature, humidity, vapour pressure, evaporation) the interpolation used was Hutchinson's spline applied to latitude, longitude, and elevation, without transforming the data. The variance estimates for these variables were calculated for each day of the year, for each station, using a five-day moving average to increase the number of data points available. Thus, the variance for a given station for all October 5s was calculated from the data for that station for Oct 3, 4, 5, 6, 7 over all years.

Rainfall data are likely to be of variable accuracy in areas where rainfall station density is low in absolute terms. Station numbers were lower in the arid interior of the country, the North, and the North West (Australian Bureau of Meteorology 2000a). The density of stations for the other weather data in Australia is much lower than the density of rainfall stations, and most sparse in the arid interior, the North West, southern Queensland (except just around the coastal strip), and northern New South Wales. The estimated values were also likely to be less accurate where station density was low relative to climate gradients (such as close to the coast and in areas of topographic complexity). There is likely to be some conservatism in the estimation of the rainfall data: the interpolation accurately estimates low rainfall amounts, but tends to slightly underestimate large rainfall amounts (Jeffrey et al. 2001). This is an artefact of all interpolation methods, which generally produce a value between the input values, rather than outside the range. The interpolation error increases in summer months due to the onset of convective rain storms. This error is exacerbated by the fact that the density of rainfall
stations is lower in the northern part of Australia, which experiences monsoonal-type rainfall (Jeffrey et al. 2001). This is not expected to have caused bias in the study areas, which were both located in the southern part of the country.

Wind speed data were obtained, but the quality was considered very poor (Keith Moodie pers. comm. 1999). Wind stations are sparse in Australia, and wind information was not collected continuously over the study period. In addition, the measurement quality was variable. Some stations have old instruments that are prone to inaccuracy (Keith Moodie pers. comm. 1999). Other stations replaced the less-accurate instruments during the study period, resulting in an (unquantifiable) recording bias. For these reasons, wind data were not used. Cloud cover could potentially be used as a proxy indicator of rainfall and temperature. However, continuous cloud cover data were not available for the whole period or the geographic area of the study.

3.4.2.3 Long-term average climate data

I obtained long-term average monthly climate data (also as an interpolated surface) from the Queensland Department of Natural Resources. I used these to build a profile of the “normal” climate patterns of the study areas. The long-term average data were derived from recordings for the period 1957 to 1998 (42 years). Monthly values were supplied for each of the variables mentioned. Regarding the calculation of these values:

- Average values (e.g. mean minimum temperature) were calculated as the mean of the 42 daily spatial averages for the SLA.
- To protect against the undue influence of extreme values, absolute values (e.g. absolute minimum or maximum temperature) were calculated as the lowest (highest) of the 42 daily spatial averages for an SLA for the month, not as the single lowest (highest) value recorded for a pixel in an SLA for the 42 months.

3.4.3 Tidal data

Tidal data for the Southwest study region were provided by the Coastal Management Branch of the Maritime Division of Transport, Western Australia. During the study period, historic tidal data were only recorded at the location of Bunbury in the Southwest study area (see Figure 7.2, Chapter 7). Data were provided for hourly tide height (in centimetres), from which mean monthly tide heights and absolute monthly tide heights.
were derived. Missing values (of which there were very few) were managed by averaging values for the day prior to and post the missing day(s).

3.4.4 Mosquito data

Most public health regions in Australia do not conduct regular mosquito surveillance. Mosquito trapping is usually opportunistic, and there are only a few locations (such as parts of coastal Queensland, around Darwin, and parts of New South Wales and southwestern Western Australia) that have a continuous record. As the purpose of this study was to determine if climatic and environmental factors could be used to predict epidemics of RRv disease, mosquito density data were not initially considered an essential variable for analysis. No continuous trapping had been conducted in the Murray study area (where the first predictive modelling was conducted), and mosquito data were not included in that analysis.

However, feedback on the results of the modelling in the Murray (Duane Gubler pers. comm. 2001) prompted me to consider the question of how much additional sensitivity mosquito data might provide to the prediction of RRv disease epidemics (in addition to that provided by climate data). The Southwest of Western Australia was the second area in which I conducted analysis. In that area, regular mosquito trapping has occurred in some SLAs since at least 1991. *Ae. camptorynchus* is believed to be the principal RRv vector species in the Southwest study area, and evidence from observational studies suggests that the continuation of large numbers of these mosquitoes into early summer is an essential factor for an RRv disease epidemic. I obtained the mean monthly trap number of *Ae. camptorynchus* for the months of November and December from several sources:

- Dr Michael Lindsay’s PhD thesis (1995) for the study years 1990/91 through to 1993/94.
- A paper presented at the Seventh Arbovirus Research in Australia Symposium (Lindsay *et al.* 1996b) for the study year 1995/96.

Data for 1996/97 (in which two Southwest SLAs recorded epidemics) were not available.
The mosquito data had been collected in three regions only: Peel Inlet, Leschenault, and Capel-Busselton. I followed the method used by Michael Lindsay (pers. comm. 2002) for allocating SLAs in the Southwest study region to mosquito collection regions. The data from the Peel Inlet were applied to the Rockingham, Mandurah, Murray, and Waroona SLAs; those from the Leschenault to the Harvey, Dardanup, and Bunbury SLAs; and those from Capel-Busselton to the Capel and Busselton SLAs (see Figure 7.7 in Chapter 7). No mosquito data were available for the southern five SLAs in the Southwest study area.

Within each mosquito collection region, three to ten traps were checked fortnightly, and the species type and number for each trap were recorded (Lindsay 1995). The trapped areas within the regions were chosen for their proximity to both human and reservoir population sites. Prior research has shown the mean trap number of mosquitoes per region to be a reasonable estimate of mosquito activity (Lindsay 1995), even though standard errors between trap numbers may be wide (Michael Lindsay pers. comm. 2002).

### 3.4.5 Mosquito control activities

#### 3.4.5.1 Murray River

No mosquito control activities were conducted on the northern border of the Murray River during the study period. Limited interventions were conducted in all years, in the form of sporadic larviciding, in some SLAs to the south. I attempted to obtain records of the timing and extent of control from the five relevant Victorian Local Councils. Records had been kept manually in a couple of these regions, but the labour involved in accessing them was prohibitive (for both parties). I requested a summary of these data from the Victorian Health Department, but they were unable to provide these in the thesis timeframe.

#### 3.4.5.2 Southwest

Aerial spraying was conducted occasionally in the Peel Inlet and Leschenault regions during the period, but not in the Capel-Busselton region. All three regions conducted low level adulticiding from year to year during the study period (Michael Lindsay pers. comm. 2002).
3.4.6 Irrigation data

As outlined in Chapter 1, the irrigation of crops, particularly rice crops, has the potential to increase mosquito population numbers and thus confound the influence of rainfall.

- In the Murray Valley study area, agricultural industries are heavily reliant on irrigation for regular watering of crops. I found no comprehensive source of data on irrigation for the whole area. However, I did identify data that, although not regional in coverage, were useful for suggesting ways in which irrigation may act as a substitute for rainfall in dry years. The predominant irrigation method used by cropping industries along the Murray River is the release of water from the Hume Reservoir downstream throughout the growing season (typically from September to April). The Murray-Darling Basin Commission provided daily data on flow and gauge height (measured at a single point near the town of Albury downstream from the Hume Reservoir) from January 1991 to December 1999. The Commission also provided advice on the amount of flow (in mega litres per day) required to reach and maintain flood levels at different points along the river, as well as the estimated area under inundation (in hectares) that might be achieved at particular flow levels. Because the coverage was not comprehensive, these data were not included in the models that I developed in the Murray, but helped in the interpretation of the findings for those regions.

- No rice cropping occurs in the Southwest study area. There is no evidence from previous research (e.g. Lindsay 1995) that irrigation is a factor that influences mosquito breeding cycles in that study area.
Chapter 4

Epidemiology of RRv disease in Australia: 1991-1999

4.1 INTRODUCTION

This chapter describes the distribution of RRv disease notifications in Australia during the period July 1991 to June 1999. Previous reports of RRv disease incidence have also been based on notifications (made to State or national reporting systems or laboratories: Aaskov et al. 1981a, Hawkes et al. 1985, Phillips et al. 1990, Tai et al. 1993, WHO 1994, Herceg et al. 1996, Curran et al. 1997, Thomson et al. 1998, O’Brien et al. 1999, Roche et al. 2001), or have been estimated from surveys (generally self-administered: Mudge 1977, Mudge and Aaskov 1983, Condon and Rouse 1994, Westley-Wise et al. 1996). The analysis presented in this chapter is the most comprehensive to date, in that it describes the epidemiology of disease for the whole country over an eight year period. In addition to analyses within the standard political boundaries, Statistical Local Areas (SLAs) are categorised in terms of high, expected, or low rates of RRv disease. SLAs are the smallest spatial units used by the Australian Bureau of Statistics (2002a) for the routine collection of a wide range of statistical data in Australia. The influence of long-term rainfall distribution on the seasonal pattern of disease, on endemic and epidemic transmission cycles, and on average annual incidence rates is also discussed, and is compared to what has been recorded previously.
4.2 METHODS

4.2.1 Spatial analysis

The mapping of health data is not a new technique for epidemiologists. However, the development of geographical information systems (GIS) and other map-based tools (in particular remote sensing) over the past twenty years has enabled an exploration of the complexities of environmental exposures and the spatial distribution of mosquito-borne diseases on a larger scale, accompanied by more sophisticated statistical methods than were previously possible.

GIS are computer-based systems for automating, storing, analysing, and displaying mapped information and data (Burrough 1986, Maguire et al. 1991, Liebhold et al. 1993). Data can come from standard maps, aerial photographs, satellite images, or data collected on the ground (by global positioning system receivers, standard monitoring, surveys, etc.). Spatial data can be represented in two main ways: in raster format (evenly divided grid cells to display data), and in vector format (strings of X and Y coordinates, representing points, lines, or polygon features) (Vine et al. 1997).

There are three main types of spatial analysis – visualisation, exploratory data analysis, and model building (Bailey and Gatrell 1995). A GIS can be used to visualise data, and to develop hypotheses to explore cause-effect relationships (Pfeiffer 1996). The real usefulness of a GIS, however, lies in its ability to manipulate spatially-referenced data (Garner 2001). Data can be incorporated in a GIS to produce map “layers”, and two or more layers can be combined (the process of “overlaying”). Using a common georeferencing system, a new map can be derived that is a combination of the attributes of the individual layers. For example, by referencing disease data and overlaying these with images of land use and climatic regions, it is possible to examine how disease is related to features of the environment. In this study, key “themes” that were used to investigate the RRv disease and climate relationship were cases, population, rainfall, vegetation, temperature and humidity.

In one sense, a potential disadvantage of GIS lies with the ability to use them to produce high quality cartographic images. The output of a GIS is only as good as the data that it is based on, and missing or incorrect data are more difficult to identify in a map presentation than in a more traditional format (such as tables). In addition, the success of
GIS as an "enabling technology" will depend on the quality of understanding of the spatial and temporal relationship between vector and disease distribution (Rogers and Williams 1993) – in other words on the biological plausibility of the timing and mechanism of disease transmission.

GIS technology was used in the following manner in this chapter:

- The notification data were given a spatial reference, so that each record could be allocated to a location and mapped.
- Preliminary analysis (accompanied by standard statistical tools) was conducted at the national, state, and macro-climate levels to explore patterns of disease.

In Chapters 5, 6, and 7:

- Spatial macro-climate data were used to develop a geographical classification system suitable for the RRv modelling.
- GIS was used to summarise the exposure data (climate and environment factors) for the spatial units.

### 4.2.2 The base spatial unit

The National Notifiable Disease Surveillance System (NNDSS) employs the "postcode area" as the spatial identifier for all communicable disease notifications. Postcode areas are developed and maintained by Australian Post, for the purpose of mail delivery, and do not record underlying population figures. The Australian Bureau of Statistics is responsible for maintaining the Australian Standard Geographical Classification (ASGC). The ASGC is used for the collection and presentation of a broad range of social, demographic, and economic statistics, including the population census. As a classification system, it is an amalgam of the Australian political and demographic administrative boundaries.

<table>
<thead>
<tr>
<th>ASGC</th>
<th>Abbreviation</th>
<th>Number of units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census districts</td>
<td>CD</td>
<td>37 209</td>
</tr>
<tr>
<td>SLAs</td>
<td>SLA</td>
<td>1272</td>
</tr>
<tr>
<td>Statistical sub-divisions</td>
<td>SSD</td>
<td>197</td>
</tr>
<tr>
<td>Statistical divisions</td>
<td>SD</td>
<td>65</td>
</tr>
<tr>
<td>States and Territories</td>
<td>S/T</td>
<td>9 *</td>
</tr>
</tbody>
</table>

* Australia has eight States and Territories, but the ASGC system counts offshore Territories as a separate classification group.
Chapter 4 – Epidemiology of RRv disease in Australia

The ASGC has five hierarchical levels (Table 4.1, and illustrated in Figure 4.1). In this structure, census districts aggregate to form SLAs, SLAs aggregate to form Statistical Sub-divisions, and this aggregation principle continues up the remaining hierarchical levels. At each hierarchical level, the component spatial units collectively cover all of Australia, without gaps or overlaps.

In non-census years, census districts are undefined, and the main structure has only four levels of hierarchy. In these years, the SLA is the base spatial unit used to collect and disseminate statistics, other than those collected from the population censuses. SLAs are based on the boundaries of incorporated bodies of Local Governments, where these exist (Australian Bureau of Statistics 2002a).

Statistical Sub-divisions and Statistical Districts are the larger regions, and consist of one or more SLAs. Statistical Sub-divisions are defined as socially and economically homogeneous regions, under the unifying influence of one or more major towns or cities. Statistical Districts are predominantly urban areas, the boundaries of which are designed to contain the anticipated urban spread of the area for at least 20 years. They are generally defined as containing an urban population of 25 000 or more.

The SLA was chosen as the base unit for the disease data in this study because: (i) it was the smallest spatial unit for which population data were available for the whole of Australia during the study period, and (ii) it was small enough for aggregation into larger “RRv bioclimatic” regions (see Chapter 5). For the analyses in this chapter, the objective was to describe the distribution of RRv disease cases in space and time, using SLAs as the base spatial unit.

4.2.3 Georeferencing the disease data

To enable notifications of RRv disease to be displayed at the SLA level, it was necessary to allocate locations to the NNDSS data, which is reported by postcode. Georeferencing (also known as geocoding) refers to the process of assigning a spatial reference system to data, in this case X and Y coordinate points. It is the method of locating and linking cases to the underlying population at risk (Elliott et al. 1995) and to the exposure factors that will be tested. Desktop mapping software (MapInfo version 6.5 1998) was used to identify the physical centroid of each postcode area, which were then matched to the SLAs that contained them.
Figure 4.1  Main Australian political, demographic, and postal classifications (1996 boundaries).
The Northern Territory data were treated differently, as the Territory reports communicable diseases by health regions rather than by postcodes. To avoid treating the whole of the Northern Territory as one spatial unit, I obtained the health region code for each RRv disease record, and a map of the health regions (which were essentially aggregates of SLAs), from the Northern Territory Health Department. I gave each Northern Territory RRv disease record a health region number that operated as a spatial link to a digitally generated map.

Some RRv disease notifications had no postcode, and could not be georeferenced. From the dataset of 39,422 records, 1,959 records did not list a postcode. After accounting for the 1,810 records from the Northern Territory, there were 149 records without a postcode. A further 136 had invalid postcodes. In all, 285 (0.7%) notifications could not be georeferenced.

### 4.2.4 Calculating population rates

Data were checked for inaccurate codes in EPIINFO (Dean et al. 1990), and rates and expected cases were calculated in Excel (Microsoft Corporation 2000). The data were analysed within “RRv disease years” (July of one year to June of the next), to reflect the annual RRv disease epidemic curve. Notifications typically commence in the austral spring (September to November) of one year, peak in austral mid-summer to early-autumn (January to April), and drop off in the late-autumn and winter of the next year. For all analyses, the estimated “date of onset” of a case was used.

Rates were calculated after adjusting for age and sex differences within each SLA. These have been observed to independently influence the risk of infection, and the probability of disease given infection (Chapter 1). No age was recorded for 172 records, and the sex category was coded “unknown” for 787 records. In total, 38,463 records were used to determine annual age and sex standardised notification rates for the SLAs, where:

\[
\text{Incidence (SLA)} = \frac{\text{observed cases by 5-year age group and sex in each SLA}}{\text{age group and sex specific population total for each SLA}} \times 100,000
\]

Figure 4.2 illustrates the Australian population density, which is highest around the very fringes of the country (particularly the eastern seaboard) and extremely low inland. The SLA sizes broadly reflect this density distribution: SLAs are small in cities and larger in rural areas, which affords approximately comparable denominator population sizes in some parts of the country. Overall, however, SLA population sizes vary substantially. Of
the 1272 SLAs in the 1996 ASGC, the median number had 6,110 people (with a range of 3 to 250,496).

A population census is conducted every five years in Australia (1991, 1996, 2001, etc.). Estimated population totals were nominally available for all years in the study period. In practice, this was not the case. Population figures for the inter-censal years were not available for some 427 (~33%) of the SLAs, due to boundary changes and subsequent problems with estimating continuous denominator populations. Accordingly, the 1996 census figures were used to calculate expected cases and observed rates of infection for each SLA, being the census closest to mid-point for the period of the study (i.e., 1991-1999).

4.2.5 Calculating expected cases

To test whether demographic factors were responsible for differences in notifications between SLAs, the observed number of cases was compared with the expected number of cases based on population size, after standardising for the age and sex structure of the population. The number of cases of RRv disease in a region follows a Poisson distribution (Checkoway et al. 1989). The annual (i.e., July to June) expected number of cases for each SLA was calculated by:

$$E_i = \frac{\text{sum of } (c_{x_i} \times r_{x_i})}{100,000}$$

where,

- $E_i$ = expected number of cases in an SLA per year by 5-year age category and sex,
- $c_{x_i}$ = age group and sex specific infection rate per year in the national population,
- $r_{x_i}$ = age group and sex specific annual population estimate within an SLA.

$c_{x_{ij}}$ was calculated by:

$$c_{x_{ij}} = \frac{\text{observed cases by 5-year age group and sex (national)}}{\text{population by age group and sex (national)}} \times 100,000$$

$c_{x_{ij}}$ was calculated based on 1996 population data (all decimals were rounded up).

The relative level of RRv disease activity between SLAs was measured by the chi statistic, following Garner and others (1999). $Y_i$, the variation between expected and observed cases per annum in regions was calculated by:

$$Y_i = \frac{(O_i - E_i)}{\sqrt{E_i}}$$

where,

- $O_i$ = the observed number of cases in the ith region,
- $E_i$ = the expected number of cases in the ith region (see calculation above).
Figure 4.2  Australian population density by Statistical Local Area, and State and Territory boundaries.
The expected number of cases was assumed to be proportional to population size. Yi has a mean of 0 and a variance of 1. Calculating the chi statistic as \((O_i - E_i)/\sqrt{E_i}\) gives numbers that show the direction of deviation from the expected value. Positive values in excess of two standard deviations were taken to indicate greater than expected cases during the period, and negative values less than two standard deviations indicated fewer cases than expected.

### 4.2.6 Rainfall zones

In addition to an analysis of notifications by State and Territory, I also conducted analyses within rainfall zones (Australian Bureau of Meteorology 2000a), based on aggregates of SLAs. The purpose of this assessment was to determine whether long-term average rainfall boundaries could better explain the pattern of notification rates, and their seasonal distribution. A map of rainfall zones was overlayed by a map of SLAs, thus enabling each SLA to be allocated to a rainfall zone (see Chapter 5 for methodological details and a map of the rainfall zones). The definition of the rainfall zones is given in Table 4.2.

#### Table 4.2 The major Australian rainfall zones.*

<table>
<thead>
<tr>
<th>Rainfall zone</th>
<th>Zone definition **</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant winter</td>
<td>&gt; 350 mm per year (seasonal ratio &gt; 3)</td>
</tr>
<tr>
<td>Mostly winter</td>
<td>&gt; 350 mm per year (3 &gt; seasonal ratio &gt; 1.3)</td>
</tr>
<tr>
<td>Uniform</td>
<td>&gt; 350 mm per year (seasonal ratio &lt; 1.3)</td>
</tr>
<tr>
<td>Arid, mostly uniform</td>
<td>≤ 350 mm per year (seasonal ratio &lt; 1.3)</td>
</tr>
<tr>
<td>Arid, mostly summer</td>
<td>≤ 350 mm per year (3 &gt; seasonal ratio &gt; 1.3)</td>
</tr>
<tr>
<td>Mostly summer</td>
<td>≥ 350 mm per year (3 &gt; seasonal ratio &gt; 1.3)</td>
</tr>
<tr>
<td>Dominant summer</td>
<td>&gt; 350 mm per year (seasonal ratio &gt; 3)</td>
</tr>
</tbody>
</table>

* Zone definitions are from the "Seasonal Rainfall Zones", Australian Climate Atlas, Australian Bureau of Meteorology, 1975.

** The seasonal ratio is the ratio (greater : lesser) of the median rainfall for the austral summer (November to April) and austral winter (May to October).
4.3 RESULTS

4.3.1 Notification rates

From July 1991 to June 1999, there were 39,422 notifications of RRv disease to the NNDSS (Communicable Diseases Network Australia 1999). For the reasons discussed in Chapter 3, this is probably an under-estimate due to under-reporting. The average annual number of notifications during this period was 4,928, and notifications peaked in 1995/96 at 7,653 (Figure 4.3).

Results of the spatial distribution of RRv disease are presented by State and Territory, for comparison with previous studies (Figure 4.2 shows the State and Territory boundaries). Table 4.3 gives the notification numbers and rates per 100,000 by State and Territory. The average annual notification rate was 25.9 per 100,000 residents, varying from 13.6 per 100,000 in 1994/95 to 40.2 per 100,000 in 1995/96. The greatest total number of cases occurred in Queensland (58.2%), with New South Wales (14.3%) and Western Australia (10.5%) reporting the next highest numbers. After adjusting for the size of the population, people living in the warm northern half of the continent were in general at much higher risk of infection than elsewhere. Although the percentage contribution of notifications from the Northern Territory for the period was low (4.6%), the average annual rate was 110.3 per 100,000, and the highest annual rate was in 1994/95 (179.9 per 100,000). The State with the second highest average annual rate was Queensland (80.5 per 100,000). The lowest average annual notification rates were in the southern States of Tasmania (5 per 100,000) and Victoria (7.9 per 100,000).

The south-eastern States of South Australia and Victoria showed the same inter-annual pattern as each other. Epidemics occurred in 1992/93 (54.4 per 100,000 and 26.1 per 100,000 respectively) and 1996/97 (43.4 per 100,000 and 21.6 per 100,000 respectively) in both States, with very low rates in intervening years. The lowest annual rate in these States was in 1994/95 (1.2 per 100,000 and 0.6 per 100,000). Notification rates in New South Wales were less than 10 per 100,000 in most years. This State shared peak notification years with Queensland in 1995/96 (15.5 per 100,000), and South Australia and Victoria in 1996/97 (24.7 per 100,000). RRv disease was not notified in the southern State of Tasmania.

\[1\] Notifications in the Australian Capital Territory were very low (29 for the 8 year period). The current understanding is that no RRv vectors inhabit the Territory (Ian Marshall pers. comm. 2000), and these notifications of RRv disease in all probability reflect infection that occurred elsewhere. For these reasons, notifications for the Territory are not referred to in the discussion of the results.
until 1994/95, and the rates since then have remained low (ranging from 1.6 per 100 000 in 1997/98 to a peak of 14.8 per 100 000 in 1998/99).

**Figure 4.3** RRv disease notifications in Australia by year (July 1991 to June 1999).

![Graph showing RRv disease notifications in Australia by year](image)

### 4.3.2 Age and sex distribution

In general, people were at increased risk of infection in their middle (active) years of life. Notifications were highest in the 35 to 39 year age group (14.2%). This is similar to the peak ages reported in the New South Wales outbreak (n=920: Hawkes *et al.* 1985), and slightly older than the peak age of cases reported for Australia in 1980-81 (n=528: Mudge and Aaskov 1983).

The median age of disease notification for the country was 39 years (Table 4.4), with a similar pattern at the State level (a median age range of 39 to 42 years). These medians were one to five years more than the median age of the State and Territory populations, which suggests that the average age of notification is partly a function of the average population size in each State and Territory. People living in the Northern Territory had the lowest median age of notification (36 years), reflecting the lower average age of Territorians, younger than people in other States by five to eight years. A similarly low mean of 34 years was recorded among cases in an outbreak in the Northern Territory in 1990-91 (n=368: Tai *et al.* 1993).
Table 4.3

RRv disease notifications and rates per 100,000 per annum, and average annual notification rates, by State and Territory of notification (July 1991 to June 1999).

<table>
<thead>
<tr>
<th>Year</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
<th>Case Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>92/93</td>
<td>150.8</td>
<td>177.7</td>
<td>10.6</td>
<td>8.2</td>
<td>3.5</td>
<td>1.6</td>
<td>2.1</td>
<td>1.2</td>
<td>0.9</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>93/94</td>
<td>140.0</td>
<td>173.3</td>
<td>9.9</td>
<td>6.2</td>
<td>2.9</td>
<td>1.4</td>
<td>1.8</td>
<td>1.1</td>
<td>0.8</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>94/95</td>
<td>130.2</td>
<td>170.6</td>
<td>9.1</td>
<td>5.6</td>
<td>2.6</td>
<td>1.3</td>
<td>1.7</td>
<td>1.0</td>
<td>0.7</td>
<td>0.4</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>95/96</td>
<td>120.4</td>
<td>167.3</td>
<td>8.3</td>
<td>5.3</td>
<td>2.5</td>
<td>1.2</td>
<td>1.6</td>
<td>0.9</td>
<td>0.6</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>96/97</td>
<td>110.6</td>
<td>164.0</td>
<td>7.6</td>
<td>5.0</td>
<td>2.4</td>
<td>1.1</td>
<td>1.5</td>
<td>0.8</td>
<td>0.5</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>97/98</td>
<td>100.8</td>
<td>160.9</td>
<td>7.0</td>
<td>4.7</td>
<td>2.3</td>
<td>1.1</td>
<td>1.4</td>
<td>0.7</td>
<td>0.5</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>98/99</td>
<td>91.0</td>
<td>157.1</td>
<td>6.4</td>
<td>4.5</td>
<td>2.2</td>
<td>1.0</td>
<td>1.3</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

*NB: No data were collected in Tasmania for the first three years of the study period.*
Table 4.4 RRv disease notifications, median age of notification, and median age of the general population, by State and Territory (July 1991 to June 1999).

<table>
<thead>
<tr>
<th>State</th>
<th>Total notifications</th>
<th>Notification median age</th>
<th>Population median age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern Territory</td>
<td>1820</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>Queensland</td>
<td>22999</td>
<td>39</td>
<td>35</td>
</tr>
<tr>
<td>South Australia</td>
<td>1725</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>New South Wales</td>
<td>5663</td>
<td>40</td>
<td>36</td>
</tr>
<tr>
<td>Victoria</td>
<td>2977</td>
<td>40</td>
<td>36</td>
</tr>
<tr>
<td>Western Australia</td>
<td>4136</td>
<td>40</td>
<td>35</td>
</tr>
<tr>
<td>Tasmania</td>
<td>201</td>
<td>42</td>
<td>38</td>
</tr>
<tr>
<td>National</td>
<td>39422</td>
<td>39</td>
<td>36</td>
</tr>
</tbody>
</table>

The age structure of RRv disease cases cannot be directly related to the broader demographic, however. At the national level, 70% of cases occurred in people aged between 25 and 54 years. By comparison, only 44% of the total 1996 population were within the same age groups. The notification rate was more than 43 per 100 000 for people aged 35-54 years, and peaked at 49 per 100 000 (35-39 years). A significant dip can be noted in the age group notification rate curve for the 30-35 year age group ($\chi^2=627$, $p<0.001$, 95% CI) – illustrated in Figure 4.4.

Figure 4.4 Average annual RRv disease notification rates in Australia, by five year age group (July 1991 to June 1999).
The younger and older age groups experienced lower rates of infection. Thirteen percent of the general population was less than 10 years old, yet only 1% of notifications were reported in children of those ages (a rate of 2 per 100,000). As noted in Chapter 1, RRv disease in children is relatively uncommon (Fuller and Warner 1957, Wilson 1957, Seglenieks and Moore 1974, Mudge 1977, Mudge and Aaskov 1983, Hawkes et al. 1985, Tai et al. 1993). The explanations offered for this include less frequent exposure to mosquitoes (Hawkes et al. 1985) and a different clinical response to adults (for a review, see Harley et al. 2001). The lower rates of infection observed in people over 80 years (less than 10 per 100,000) has been assumed to be a function both of higher levels of immunity (Boughton et al. 1984) and of decreased exposure to mosquitoes.

Slightly more notifications were reported among women (19,607) than men (19,023) for the period, although this difference was not significant ($\chi^2=0.16$, $p<0.68$, 95% CI) (see Figure 4.5).

**Figure 4.5** National RRv disease notification rates per 100,000 by sex, for the years 1991/92 to 1998/99.

At the state level, several differences in the male to female ratio of RRv disease notifications were evident (Table 4.5). The 3% greater number of cases among women in New South Wales was not significant ($\chi^2=0.011$, $p<0.92$), and reflects the underlying population distribution. After accounting for the broader sex distribution, significantly more notifications were reported for females in Queensland (9%, $\chi^2=32.84$, $p<0.001$) and Victoria (16%, $\chi^2=7.6$, $p<0.005$). Significantly more were reported for males in Western
Australia (20%, $\chi^2=56.67, P<0.001$), the Northern Territory (17%, $\chi^2=7.2, P<0.007$), and South Australia (9%, $\chi^2=6.81, P<0.009$).

Table 4.5 National RRv disease notifications for males and females, sex ratio, general population ratio, and sex status unknown, by State and Territory (July 1991 to June 1999).

<table>
<thead>
<tr>
<th>State*</th>
<th>Males</th>
<th>Females</th>
<th>Sex ratio</th>
<th>General population</th>
<th>Sex unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
<td>%</td>
<td>cases</td>
<td>(M:F)</td>
<td>cases</td>
</tr>
<tr>
<td>NSW</td>
<td>2760</td>
<td>49%</td>
<td>2836</td>
<td>1: 1.03</td>
<td>32</td>
</tr>
<tr>
<td>NT</td>
<td>987</td>
<td>55%</td>
<td>821</td>
<td>1: 0.83 **</td>
<td>2</td>
</tr>
<tr>
<td>Qld</td>
<td>10718</td>
<td>47%</td>
<td>11726</td>
<td>1: 1.09 **</td>
<td>499</td>
</tr>
<tr>
<td>SA</td>
<td>899</td>
<td>52%</td>
<td>822</td>
<td>1: 0.91 **</td>
<td>49</td>
</tr>
<tr>
<td>Vic.</td>
<td>1262</td>
<td>43%</td>
<td>1459</td>
<td>1: 1.16 **</td>
<td>247</td>
</tr>
<tr>
<td>Tas.</td>
<td>90</td>
<td>45%</td>
<td>108</td>
<td>1: 1.20</td>
<td>-</td>
</tr>
<tr>
<td>WA</td>
<td>2294</td>
<td>56%</td>
<td>1824</td>
<td>1: 0.80 **</td>
<td>5</td>
</tr>
<tr>
<td>National</td>
<td>19010</td>
<td>48%</td>
<td>19596</td>
<td>1: 1.03</td>
<td>787</td>
</tr>
</tbody>
</table>

* NSW= New South Wales, NT= Northern Territory, Qld= Queensland, SA= South Australia, Vic.= Victoria, Tas.= Tasmania, WA= Western Australia.
** Significant at P<0.005, 95% confidence interval for the difference of two proportions.

Some 2% of all notifications were not coded for sex. The only States to report large numbers of “unknown” for the sex category were Victoria (247, 8.3% of State total) and Queensland (499, 2.2% of State total). There was no way of determining if there was differential reporting by sex for this category.

4.3.3 Reporting period

I calculated the period between the estimated onset of symptoms for each RRv disease notification and the date of report for each record. Results are presented in seven classes: 0 days, 1-7 days, 8-14 days, 15-28 days, 1-3 months, 3-12 months, and >12 months. Table 4.6 reports these results by State and Territory.
Table 4.6  

<table>
<thead>
<tr>
<th>States*</th>
<th>0 days**</th>
<th>1-7 d</th>
<th>8-14 d</th>
<th>15-28 d</th>
<th>1-3mth</th>
<th>3-12 mth</th>
<th>&gt;12mth</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>(d)</td>
<td>(d)</td>
</tr>
<tr>
<td>NSW</td>
<td>6</td>
<td>37</td>
<td>22</td>
<td>21</td>
<td>11</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>40.3</td>
</tr>
<tr>
<td>NT</td>
<td>2</td>
<td>31</td>
<td>28</td>
<td>24</td>
<td>14</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>29.9</td>
</tr>
<tr>
<td>Qld</td>
<td>-</td>
<td>30</td>
<td>33</td>
<td>25</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td>21.8</td>
</tr>
<tr>
<td>SA</td>
<td>3</td>
<td>26</td>
<td>19</td>
<td>31</td>
<td>20</td>
<td>2</td>
<td>0</td>
<td>16</td>
<td>23.7</td>
</tr>
<tr>
<td>Tas.</td>
<td>3</td>
<td>18</td>
<td>32</td>
<td>31</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>14</td>
<td>30</td>
</tr>
<tr>
<td>Vic.</td>
<td>31</td>
<td>16</td>
<td>16</td>
<td>22</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>26</td>
</tr>
<tr>
<td>WA</td>
<td>17</td>
<td>5</td>
<td>9</td>
<td>28</td>
<td>35</td>
<td>5</td>
<td>1</td>
<td>25</td>
<td>53</td>
</tr>
<tr>
<td>Australia</td>
<td>5</td>
<td>27</td>
<td>27</td>
<td>25</td>
<td>15</td>
<td>2</td>
<td>0</td>
<td>19.2</td>
<td>30.9</td>
</tr>
</tbody>
</table>

* NSW=New South Wales, NT=Northern Territory, Qld=Queensland, SA=South Australia, Vic.=Victoria, Tas.=Tasmania, WA=Western Australia
** Defined as the "date of report" minus the "onset of symptoms" (i.e., zero days elapsed).

The mean number of days from the onset of symptoms to the record of a positive diagnosis for the country was 19.2 days. Nearly 60% of cases were reported in the first 14 days, and 85% in the first 28 days. The recall accuracy of the date of onset is likely to be reasonably high within a one month period. For 15% of people the period between onset and report was one to three months, and recall bias is likely to affect these estimates. Only 2% reported a three to twelve month gap.

There were substantial differences between the States in the reporting periods. Victoria and New South Wales recorded the lowest mean number of days between onset of symptoms and report (9 days), and Western Australia recorded the highest (25 days). The other states ranged from an average of 11 to 14 days.

4.3.4 Spatial distribution

A comparison of notification rates reported by rainfall zones (Table 4.7 below), and by SLAs, indicates that the pattern of RRv infection is more complex than can be revealed by an analysis at the State level.

4.3.4.1 By rainfall zones

The northern half of the Northern Territory and Queensland, and the upper third of Western Australia, have a "dominant summer" (i.e., monsoonal) rainfall (see the rainfall
zones figure in Chapter 5). In all years except one, the notification rate in this zone was above 130 per 100 000 (average annual rate 142 per 100 000). A peak annual notification rate of 207.2 per 100 000 was recorded in 1992/93. The south of the Northern Territory has “arid, mostly summer” rainfall conditions. The annual notification rates in this zone were higher than the national average for the period, but still much lower than in the northern part of the Territory (an average annual notification rate of 57.4 per 100 000).

<table>
<thead>
<tr>
<th>Year</th>
<th>Dominant winter</th>
<th>Mostly winter</th>
<th>Uniform</th>
<th>Arid, uniform</th>
<th>Arid, summer</th>
<th>Mostly summer</th>
<th>Dominant summer</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>32.6</td>
<td>3.2</td>
<td>1.4</td>
<td>57.3</td>
<td>62.5</td>
<td>77.9</td>
<td>131.5</td>
</tr>
<tr>
<td>92/93</td>
<td>7.7</td>
<td><strong>44.1</strong></td>
<td>4.0</td>
<td><strong>229.9</strong></td>
<td>60.5</td>
<td>23.9</td>
<td><strong>207.2</strong></td>
</tr>
<tr>
<td>93/94</td>
<td>3.2</td>
<td>1.4</td>
<td>2.0</td>
<td>7.8</td>
<td>38.4</td>
<td>72.8</td>
<td>157.4</td>
</tr>
<tr>
<td>94/95</td>
<td>4.0</td>
<td>0.8</td>
<td>1.4</td>
<td>24.8</td>
<td>85.9</td>
<td>26.9</td>
<td>145.1</td>
</tr>
<tr>
<td>95/96</td>
<td><strong>75.7</strong></td>
<td>5.7</td>
<td>5.0</td>
<td>16.9</td>
<td>45.6</td>
<td><strong>122.5</strong></td>
<td>138.7</td>
</tr>
<tr>
<td>96/97</td>
<td>22.7</td>
<td>36.1</td>
<td><strong>13.6</strong></td>
<td>186.9</td>
<td><strong>214.1</strong></td>
<td>41.1</td>
<td>170.6</td>
</tr>
<tr>
<td>97/98</td>
<td>23.5</td>
<td>7.0</td>
<td>5.7</td>
<td>22.2</td>
<td>40.4</td>
<td>38.4</td>
<td>131.7</td>
</tr>
<tr>
<td>98/99</td>
<td>9.4</td>
<td>3.5</td>
<td>6.0</td>
<td>19.5</td>
<td>78.1</td>
<td>55.4</td>
<td>53.6</td>
</tr>
<tr>
<td>Period</td>
<td>22.3</td>
<td>12.8</td>
<td>5.0</td>
<td>70.7</td>
<td>78.2</td>
<td>57.4</td>
<td>142.0</td>
</tr>
</tbody>
</table>

* The highest annual rates in each zone are bolded.

In the south-east of the continent, people at highest risk of RRv infection lived in the “arid, mostly uniform” rainfall zone, encompassing the western plains of New South Wales, and bordering the Murray Valley of inland eastern Australia. Rates in this zone were equivalent to or higher than the national average for most of the years, with two extremely high peaks (229.9 per 100 000 in 1992/93, and 186.9 per 100 000 in 1996/97). Notification rates in the “uniform” rainfall zone were much lower than the national average for all years (average annual rate of 5.0 per 100 000).

The arid rainfall zones recorded the highest annual rates in the country (229.9/100 000 in 1992/93 in the arid uniform zone, and 214.1/100 000 in 1996/97 in the arid summer zone). The average annual notification rate in both zones was lower than in the northern dominant rainfall zones, however, skewed downwards by lower RRv notification rates in between these epidemic years.
4.3.4.2 By SLA

Seventy-one percent (906) of the 1272 SLAs reported at least one case of disease over the eight year period. However, notifications were not evenly distributed: 90% were reported from only 28% (362) of the SLAs. Figure 4.6 is a map of Australian SLAs shaded according to the Chi statistic. Positive values indicate more notifications than expected based on population distribution; negative values indicate less notifications than expected; a value of zero indicates equilibrium. In 28% of SLAs the observed number of cases exceeded the expected for the period, and in 34% there were fewer cases than expected, indicating that not all cases can be explained by population effects alone.

In the Northern Territory, the risk of infection was greatest in the northern coastal areas, and somewhat reduced (although still higher than average) in Central Australia – a gradient noted by Tai and others (1993) in their report of the 1990-1991 outbreak in that Territory. In Queensland, the risk of infection was highest in the band of SLAs that runs the length of the State along the eastern border of the Great Dividing Range (Figure 4.6), where rainfall is highest. Previous studies have noted extremely high incidence rates in the central coast and Far North regions of Queensland (Aaskov et al. 1981a, Phillips et al. 1994, WHO 1994). In parts of western Queensland, although temperatures are still warm, rainfall is lower and the risk of infection was average.

The risk of infection was average for people living on the slopes of the Great Dividing Range. People were at lower than average risk along the Tablelands and the coast, south of mid-New South Wales down to Victoria and the island of Tasmania. These regions fall within the “uniform” rainfall zone. This gradient of risk – decreasing from the arid central areas towards the coast, and from the north to the south along the coast – was also apparent in the large outbreak of RRv disease across New South Wales in the summer of 1983/84 (Hawkes et al. 1985). In Western Australia, people living in the temperate coastal south-west of the State, and the Pilbara and Gascoyne regions of the arid north-west, were at high to very high risk of infection.

4.3.5 Seasonal effects

RRv disease was notified in every month of the year in all States (see Figure 4.7 of notifications by State), although it occurred more frequently in and immediately following the warmer months. For the country as a whole, RRv disease notifications commenced in November, peaked in March (25.1%), and dropped back down to low
Figure 4.6 Statistical Local Areas shaded according to the Chi statistic. A value of zero indicates equilibrium; positive values indicate more RRv disease notifications than expected on the basis of population distribution; negative values indicate less than expected.
Figure 4.7  RRv disease notifications in Australia by month and year for States and Territories (July 1991 to June 1999).
Figure 4.7 (cont’d) RRv disease notifications in Australia by month and year for States and Territories (July 1991 to June 1999).
levels in July (Figure 4.8). Fifty-two percent of notifications occurred in the austral autumn (March to May), 33.8% in austral summer (December to February), 7.6% in austral winter (June to August), and 6.2% in austral spring (September to November).

Figure 4.8  RRv disease notifications in Australia, by month of onset (July 1991 to June 1999).

There was noticeable spatial variation in the estimated season of RRv infection across the country (Table 4.8). In the south-western corner of Australia, where rain falls predominantly in winter, 60% of cases occurred in summer. In a large RRv disease outbreak in that part of the country in 1988/89, Lindsay (1989) also noted a peak of cases in January to February. In the “mostly winter” rainfall zone, there was a predominance of cases in the summer (52%), with the second largest proportion in autumn (40%). In the New South Wales-wide outbreak of 1983/84, the largest number of cases occurred in the south-western plains and the central and south-western slopes (which are within this rainfall zone), with 60% of these during summer (January to February) (Hawkes et al. 1985).

In the temperate and arid regions where rain falls, on average, uniformly throughout the year, cases were more evenly spread between summer and autumn. In areas where rain falls mostly or only in summer, the bulk of cases (~60%) occurred in autumn. RRv disease has been described as a “wet season phenomenon” in the Northern Territory (Tai et al. 1993), and previous outbreaks have peaked in late summer to autumn, following heavy summer rainfall (for example Halliday and Horan 1943, Whelan et al. 1996). A peak of cases in autumn was also noted in Queensland in 1980-81 (Mudge and Aaskov 1983).
Chapter 4 - Epidemiology of RRv disease in Australia

Table 4.8  Percentage of RRv infections each season, by rainfall zone.

<table>
<thead>
<tr>
<th>Season</th>
<th>Dominant winter (%)</th>
<th>Mostly winter (%)</th>
<th>Uniform</th>
<th>Arid (uniform) (%)</th>
<th>Arid (summer) (%)</th>
<th>Mostly summer (%)</th>
<th>Dominant summer (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter</td>
<td>5</td>
<td>2</td>
<td>9</td>
<td>5</td>
<td>14</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>Spring</td>
<td>9</td>
<td>5</td>
<td>9</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Summer</td>
<td>60</td>
<td>52</td>
<td>32</td>
<td>40</td>
<td>22</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Autumn</td>
<td>26</td>
<td>41</td>
<td>50</td>
<td>52</td>
<td>60</td>
<td>61</td>
<td>57</td>
</tr>
</tbody>
</table>

4.4 DISCUSSION

Other than annual reports of the NNDSS, which have recorded RRv disease by Statistical Divisions for individual years since 1995 (Herceg et al. 1996, Curran et al. 1997, Thomson et al. 1998, O’Brien et al. 1999, Roche et al. 2001), there have been two other national descriptions of RRv disease. The first was a self-administered survey of cases notified to the Communicable Diseases Intelligence in 1980 and 1981 (n=528, 51% response rate: Mudge and Aaskov 1983). This study did not have accurate information on the date of onset, and provided almost no information on the spatial and temporal distribution of cases. The second was a very short report to the World Health Organization (1994) of RRv disease notifications to the NNDSS for 1992. The current study represents the most comprehensive assessment of the national distribution of RRv disease, and of the influence of climate on this distribution.

4.4.1 Data quality

Several points relating to the disease and population data used for these analyses are relevant. First, the calculation of rates will have been affected to some degree by the amount of in- and out-migration in some SLAs. Human “herd immunity” to RRv infection – the proportion of susceptibles (or non-immunes) in a population (Last 2001) – decreases when (i) children who have no immunity are born into the population, or (ii) non-immune people in-migrate. Conversely, it increases as (i) the population ages (people who have been infected with RRv develop life-long immunity), or (ii) non-immune people migrate elsewhere. In- and out- migration can be extensive in inner city...
areas (Elliott et al. 1995). The rates calculated for metropolitan SLAs are likely to be affected by this factor, in addition to the other reasons already noted for suspecting the reliability of RRv notification rates in city areas. Regarding the rural areas of Australia, Haberkorn and others (Haberkorn et al. 1999) calculated changes in rural populations from 1991 to 1996. They estimated that the annual population growth over the period was 1.2 per cent for the country as a whole, with very similar growth rates for both metropolitan and non-metropolitan areas alike. Despite overall similarities, there were substantial variations between some rural areas in terms of population growth and decline. Approximately half the SLAs in the two Murray study regions (Chapter 6) showed a 0-5% population decrease during this period, and the rest showed a 0-5% population increase (Haberkorn et al. 1999). These were evenly spread between the two regions. In the Southwest study area (Chapter 7), the SLAs had between a 5 and 20% population increase during 1991 to 1996.

Second, although the postcode system is widely used as an anonymous locational case identifier, it was never developed for this purpose by the Australian Postal Service. Postcode boundaries are developed with postal, rather than data collection, requirements foremost, and do not necessarily mirror ASGC boundaries. In addition, the SLA classification system can change from year to year, depending on amalgamation or disaggregation of Local Government shires. Therefore, some inaccuracies may have occurred in the process of geocoding cases, due to minor concordance problems between the postcode and SLA boundaries, particularly in central Australia where both regions are large. 3

Third, the estimates for the variation in reporting times between States should be treated circumspectly. The date of onset of symptoms (generally made by the patient with the doctor) can be an estimate only, and there is no information available to validate this accuracy. If the “date of onset” box had not been completed, it is common practice for laboratory personnel to write in the date for when the positive diagnosis was reported by the laboratory (Paul Roach pers. comm. 2002). Thus, when the “date of onset” is the same as the “date of report” (clearly impossible) this can be read as meaning “date of onset unknown” (Paul Roach pers. comm. 2002). There were 2035 (5.2%) records during the period with zero lag between the date of onset and the date of report. These were not evenly distributed: 31% of Victorian records and 17% of Western Australian records fell into this category, whereas Queensland had none. Not surprisingly, these records were

3 Postcodes that straddled two SLAs were apportioned to the one that had the greater land area.
more common in years when the highest case numbers were reported. This suggests that missing data may have been related to general practitioner overload. The "date of report" can be variously interpreted. It can mean: (i) the date a positive diagnosis was made at a laboratory, (ii) the date a case was reported by a laboratory to a State Health Department, (iii) the date a case was reported by a general practitioner to a State Health Department, or (iv) the date a report was received by a State Health Department. To complicate matters, diagnoses can be sent by mail or electronically (although increasingly they are the latter). The mail system introduces a further element of delay. Consequently, there is likely to be intra- and inter-State differences in the length of the reporting period due to these factors.

The time lapse between date of onset of symptoms and date of report also has some bearing on the accuracy of RRv disease diagnoses. In 41% of cases, the date of report was more than fourteen days after the onset of symptoms. The acute phase of illness for most arboviruses lasts five to fifteen days (Shope and Meegan 1997), with a mean of between seven and nine days noted for RRv (Fraser and Cunningham 1980). The time it takes to collect blood from a patient, test it, and record the result, has not been studied. If it takes more than seven days, this suggests that a sizeable proportion of these cases would have to have been presumptive, as sera could not have been provided during the acute phase of the disease.

4.4.2 Sex differences in notifications

Previous studies that have reported sex-related differences in RRv disease may have inaccuracies due to a response bias (Aaskov et al. 1981a, Mudge and Aaskov 1983, Selden and Cameron 1996, Westley-Wise et al. 1996), or have not reported the underlying population sex ratio for comparison in the study population (Aaskov et al. 1981a, Fraser et al. 1986, Tai et al. 1993). The study by Hawkes and others (1985) reported an even ratio of males and females infected in New South Wales, after adjusting for the underlying population (the same result found in the findings reported here). In the absence of evidence of a true difference in receptivity to RRv infection between men and women, it seems reasonable to conclude that the sex differences in notifications reported in several States during the period were due to gender-related factors that affected exposure (such as occupation choice or opportunity, recreational activities, or attitudes to self-protection) or disease reporting. Well-conducted qualitative research would be required to provide more State-specific information on these gender-related risk factors.
4.4.3 Patterns of RRv distribution in Australia

Australia is a large landmass, and across the continent there are extreme climate changes, from monsoonal tropics in the north to arid desert inland, and from temperate coastal regions and sub-tropics in the east and south-west to cool temperate and alpine regions in the south-east. Analysis of a disease such as RRv at the national scale will be too coarse to show meaningful patterns. The occurrence varies enormously at the State and Territory levels. The Northern Territory and Queensland had the highest rates, and Tasmania the lowest. State boundaries are also clearly not a rational basis for exploring patterns for a disease like this, as others have found. For example, Harley and others (1996) concluded that no relationship could be detected between the SOI and RRv disease notifications at either the national or State levels, even though the SOI has been correlated with outbreaks of RRv disease at finer scales (Tong et al. 1998).

Rainfall zones have been used previously to describe patterns of RRv disease in New South Wales (Boughton et al. 1984, Hawkes et al. 1985). This approach has an intuitive logic, as rainfall clearly influences RRv disease distribution (Chapter 1). The broad patterns (i.e., epidemic/endemic, season of onset, average annual notification rate) were better reflected than at the State level. New South Wales, for example, is dominated by four different rainfall patterns, and the peaks in that State’s annual notification rates for the period were influenced by epidemics that occurred in either the semi-arid south-west of the State around the Murray River (shared by South Australia and Victoria), or the sub-tropical northern areas (bordering Queensland). The average risk for RRv disease was higher in the “dominant summer” rainfall zone of north Australia than in the “mostly summer” and “arid summer” rainfall zones in the middle of the country. This finding is consistent with previous reports of RRv disease incidence, which have noted a north-south decline in incidence rates from the Northern Territory to central Australia (Tai et al. 1993) and from northern Queensland to southern Queensland (Phillips et al. 1994, WHO 1994).

Inter-annual rainfall variability strongly influences annual infection rates (Chapter 1). Although the average annual rates were not the highest, the highest annual notification rates for the period were recorded in the two arid rainfall zones. Mean rainfall in these zones is very low, but highly variable in distribution, with a cycle of floods and droughts. In the Murray-Darling Basin (bordering on the arid zone in south-eastern Australia) over 30,000 wetlands (Murray-Darling Basin Commission 2001) provide breeding habitat for
mosquitoes. Most tributaries carry water only at times of flood and are dry the rest of the year. In these regions, the virus survives long periods of drought in desiccation resistant eggs, with sudden peaks of activity after heavy rains (Lindsay et al. 1993b). These conditions are conducive to an epidemic transmission cycle (evident in these findings), as vertebrate host immunity has time to increase substantially in the years between heavy rainfalls. Historically, the majority of RRv disease outbreaks around the Murray River have been associated with abnormally heavy rainfall preceding and during the outbreak (for example Nimmo 1928, Anderson and French 1957, Mudge and Aaskov 1983, Hawkes et al. 1985), and outbreaks in the north-western arid zone between 1984 and 1993 generally followed heavy autumn or winter rains (Lindsay et al. 1993b).

The analysis at the rainfall zone level still lacked the detail evident at the lower spatial resolution of SLA. The analysis of expected and observed cases showed that RRv disease risk can vary widely within the same rainfall zone, and even between co-located SLAs. The lower risk of infection observed in the Tablelands of New South Wales (also noted by Boughton et al. 1984) is probably due to the lower temperatures and humidity caused by altitude and a shortened transmission season, than to rainfall patterns alone. The steady geographic decline in risk along the eastern coast of Australia reflects the transition from the northern-most tropical regions (where mean temperatures above 21°C are sustained throughout the year, enabling continuous circulation of the virus), to the sub-tropical regions of south-eastern Queensland (where there is a dip in temperatures during winter), through to the temperate regions of mid-New South Wales down to Tasmania. This north-south gradient of RRv incidence has also been noted in the geographically confined limits of the North Coast of New South Wales (Westley-Wise et al. 1996).

Variations within rainfall zones can be traced to human as well as weather and environmental factors. Parts of subtropical Australia have the highest population growth rates in the country. The immigration of susceptible people from southern Australia, where there is less RRv activity and a lower population seroprevalence, is believed to have resulted in certain areas of south-eastern Queensland having a higher incidence of RRv disease and infection during the period (Ryan et al. 1999).

4.4.4 An increase in cases of RRv disease?

As discussed in Chapter 1, human contact with the RRv natural cycle has increased in many parts of Australia over the past 10-20 years due to: (i) population sprawl, and the
fashion for residential developments in salt-marsh habitats or the establishment of canal settlements (Tai et al. 1993, Russell 1994, Finn 1995, Lindsay et al. 1996a, Whelan et al. 1996), (ii) the movement of non-immune tourists (Weinstein 1997), workers, or residents to northern localities (Russell 1995), (iii) increases in mosquito populations from irrigated agriculture or wetland construction in otherwise arid inland areas (Mackenzie et al. 1994c, Russell 1995), or (iv) changes in land use and tree clearing providing new habitats for mosquito breeding (Mackenzie et al. 1994b).

Russell (1999) has suggested that notifications of RRv disease cases have increased in many parts of Australia over the past two decades. In this eight-year period there was some variation in annual case numbers in most States and rainfall zones, and the 1996/97 peak of cases was a notable feature. Overall, though, the results for this period do not show much evidence for an increase. The length of the time series since national reporting commenced (i.e., 1991) is still too short to determine whether the large case numbers and epidemic spikes recorded in some parts of Australia are due to natural fluctuations in RRv transmission cycles, relating to interannual variation in regional weather conditions, or to a real increase in case numbers. The influence of improved awareness of the disease, more sensitive diagnostic techniques, and better coordination of surveillance over the past decade will have resulted in additional notifications during this period than in the years prior to this.

In terms of future prospects, there is a high probability that the number of people infected with RRv will increase relative to the study period, as there is no indication that the factors currently intensifying the human interface with the RRv natural cycle will abate. Climate change is expected to affect mosquito distribution and abundance (and hence human cases of disease), although knowledge of RRv and the ecology of its vectors is still too limited to be able to make accurate statements about how this will change disease incidence (Russell 1998b).

4.5 Conclusions

RRv disease caused substantial morbidity in Australia during the study period. The disease primarily affects adults in their middle years. The sex differences in notifications in some States (more males in Western Australia, the Northern Territory and South Australia, and more females in Queensland and Victoria) cannot be adequately explained.
by these data or previous studies. Further research is needed to examine the likelihood of gender-related risk factors.

There was a strong seasonal and geographic trend in RRv disease occurrence. The northern warmer and wetter regions were generally at the greatest average annual risk. The highest annual notification rates for the period were reported in arid zones, however, suggesting that seasonal climate variation and vertebrate host immunity are powerful determinants of intense epidemics of disease. Although rainfall zones are adequate for reporting some of the broad patterns of RRv disease cycles and incidence, they are clearly not capable of describing interannual and regional patterns. SLAs are based on administrative boundaries – artificial for the purposes of explaining disease distribution. However, they are at a much finer resolution than State and Territory boundaries or climate zones. Additionally, population data are also available at this level. The next chapter looks at how the climate data can be aggregated into regions that are broadly similar, for the purpose of modelling the relationship between climate and RRv disease epidemics.
Chapter 5
Case study methods

5.1 INTRODUCTION

This chapter outlines the design and methods for the analytical studies that are reported in Chapters 6 and 7. To recap from Chapter 2, the aim of the analytical studies was to determine if climatic and environmental factors can be used to predict epidemics of RRv disease with sufficient accuracy and timeliness to be used for public health surveillance. The relevant study objectives were:

- to develop a geographical classification system based on climatic factors that influence the ecology of RRv vectors,
- to identify appropriate outcome variables for prediction, and
- to develop predictive models for the probability of RRv disease occurrence in two different regions of Australia.

The first part of this chapter describes the process for identifying a rational geographical classification system for RRv in Australia, and the steps that were used to develop the "RRv bioclimatic regions". The second part discusses appropriate outcome variables for public health purposes. The third part focuses on predictive modelling, the criteria used to select the two case study areas (the Murray and the Southwest), validation of the models, and the management of confounding variables.
5.2 DEVELOPING AN RRV GEOGRAPHICAL CLASSIFICATION

The aim was to divide Australia into a series of rational zones or regions that reflect the ecology of RRV. Because disease rate data were available at the Statistical Local Area (SLA) level, this necessarily involving grouping SLAs into regions that reflected similar RRV ecology.

“Environmental classification” has been defined as a procedure for grouping spatial units into groups, on the basis of similar environmental attributes (Thackway and Creswell 1992). The purpose of a classification system should determine the appropriate number and size of regions, the data sets needed to generate them (including the type and number of attributes), and the scale of the data available (Thackway and Creswell 1992). The purpose of this study suggested two main criteria for selecting the attributes. First, the attributes should influence the ecology of the mosquito (i.e., rainfall, temperature, humidity, vegetation). Second, to avoid producing a classification system with lots of small, disparate regions, few attributes (and less detail in the classes of each attribute) would be preferable to many.

5.2.1 Scale of analysis

A major factor influencing spatial analysis is the geographical scale at which the data will be analysed (Pfeiffer 1996). In the context of this study, “scale” refers to the spatial resolution at which subsequent predictive modelling was conducted.

A public health study of RRV disease and climate could be conducted at many levels, and it is worthwhile considering the benefits of focusing research to different public health jurisdictions. The Australian political structure consists of a Federal Government and eight States and Territories. The management of mosquito-borne diseases is the responsibility of State and Territory Governments, in conjunction with Local Governments. The Commonwealth Government supports the surveillance and control of mosquito-borne diseases through quarantine. It also has a role in policy development, and the provision of national arbovirus disease statistics. Thus, while analysis at the national level could be used to drive national mosquito-borne disease strategy, the opportunities for intervention are limited. In addition, given the complexity of RRV cycles across Australia, it is unlikely that useful predictive models could be developed at this scale.
Funding, policy, and decision-making occur at the State level, and prediction at these levels could be useful. However, Chapters 1 and 4 have presented evidence that shows how RRv transmission patterns differ greatly across the continent, obviously independent of political boundaries. Observations of vector and virus behaviour support the opinion that different RRv ecologies exist (Russell 1994). Accordingly, development of predictive models at the continental or state level was not considered to be a biologically appropriate approach.

On the other hand, considering the aim of this study, I did not consider it was desirable to develop predictive models at fine local scales of, say, 20-50 km². Information on weather conditions (and mosquito trapping) at this scale may be more accurate than at the regional level. However, the technical skill and resources involved in data collection and the production of models at this scale would be beyond the reach of most public health authorities. The comparative rareness of RRv disease, and the low density of people in non-metropolitan Australian areas at this scale, would not have provided sufficient statistical power to determine a significant association between the exposures and disease. The cost-effectiveness of focusing at the local level is also dubious for a disease such as RRv disease (discussed in more detail in Chapter 8). Further, the quality of the disease data (i.e., the reporting by place of residence rather than place of infection) was not adequate to support analyses at this scale.

For these several reasons, analyses at some regional level was considered appropriate. Unless otherwise noted, the term “region” in this thesis refers to a geographical area that has broadly similar macroclimatic characteristics. As discussed in Chapter 3, the base spatial unit for the disease and exposure data was the SLA (the smallest unit for which denominator population data were available during the study period). This enabled the calculation of disease rates for comparison between regions. SLA boundaries do not generally have any relationship to climatic boundaries (although they occasionally approximate water catchments). They are, in general, small enough to permit grouping into regions that are climatically similar with respect to factors that determine RRv vector distribution.
Australia has diverse kinds of natural ecosystems (a complex of terrain, weather, soils, flora and fauna) as well as human modifications to natural ones (such as settlements, agriculture, irrigated land). Several ways of segmenting the continent into regions have been developed, depending on the intended purpose of the classification.

I conducted a search of relevant texts and government reports to see whether an existing regionalisation system could be used for this study. I reviewed systems that had been developed in the fields of geography, meteorology, and entomology. I also reviewed the RRv disease literature to identify regional zones that had been used in related studies. From this search, I found two existing geographical classification systems with possibilities for this study: (i) rainfall districts, and (ii) conservation districts. The merits and limitations of each are discussed below.

Rainfall districts (Bureau of Meteorology)

The Bureau of Meteorology has divided the Australian continent into 99 physical regions, broadly based on the similarity of rainfall pattern (Australian Bureau of Meteorology 2000b). The boundaries of these districts were not explicitly based on precipitation values per se, but were decided upon after discussion between State Governments, climatologists, and river catchment programs in the early 1930’s. Few changes to the district boundaries have been made since then. The rainfall districts reportedly have no relationship with census districts, postal districts, water catchments, or shire boundaries (Australian Bureau of Meteorology 2000b), however they clearly follow (at least) major State boundaries. These districts are not evenly distributed among States, and there is a far larger number in New South Wales (n=30) than in other States (e.g. Queensland = 19, Northern Territory = 2, Western Australia = 13).

Rainfall districts have been previously used to describe RRv disease patterns (Boughton et al. 1984, Hawkes et al. 1985). As I argued in Chapter 4, however, rainfall represents only one of the weather factors that determines the timing and success of the mosquito lifecycle. Rainfall is overwhelmingly associated with mosquito breeding and abundance (especially in the tropical zone). However, temperature and humidity are also essential elements dictating the timing and duration of transmission.
Conservation regions (Interim Biogeographic Regionalisation for Australia)

Work in environmental sectors of Australia has focused on the development of a national system of standardised regionalisations for conservation planning and sustainable resource management. The Integrated Biogeographical Regionalisation for Australia (IBRA) was developed for the Australian Nature Conservation Agency, and seeks to describe the dominant landscape scale features of the country (Thackway and Creswell 1992). The major attributes used to delineate the IBRA boundaries are climate, lithology/geology, landform, vegetation, flora and fauna, and land use – factors known to influence biological productivity at continental and regional scales (Nix et al. 1988). Many of these attributes strongly determine regional vegetation coverage and, by extension, would influence mosquito ecosystems.

Several factors argued against the use of the IBRA system for this study. First, there was no research to suggest how three of the IBRA attributes (flora and fauna, lithology/geology, and landform) would impact on mosquito breeding. The extent to which these attributes dominated the classification of the IBRA regions was also not clear. Second, the availability of the system (and access for outsiders) in the medium to long-term was uncertain. Third, there was no information on how the classes for each attribute were defined. The consistency and completeness of the data for each attribute varied between different States (Thackway and Creswell 1992). Fourth, there were 80 separate regions, which seemed overly complex for this study.

In summary, neither of these two classification systems was considered appropriate for the purposes of this study. Consequently, I purpose-built a geographical classification system based on aggregates of SLAs, using long-term average climate zones.

5.2.3 Steps in building the RRv regional system

The steps for developing the RRv regional classification system were as follows:

1. Obtain maps of the long-term average climate and vegetation zones of Australia for the major climatic factors that determine mosquito breeding.
2. Overlay the SLA boundaries, and assign each one to a climate zone.
3. Combine the climate maps to produce regions, alike in the biological and climatic attributes (“RRv bioclimatic regions”).

These steps are expanded upon below.
5.2.3.1 Step 1 - Long-term average climate maps

The dominant climate phenomena that limit the distribution of the Ross River vectors are rainfall, temperature, and humidity (Chapter 1). Vegetation can provide a convenient proxy of both rainfall and temperature, amongst other climatic and geological features. I obtained maps of the long-term average rainfall zones, the long-term average (combined) temperature and humidity zones, and the Köppen classification of Australian vegetation (modified version) from the Australian Bureau of Meteorology (2000a). All three maps were available free via the World Wide Web, and are used as a standard by the Bureau and others throughout Australia.

Rainfall

This map is based on the Seasonal Rainfall Zones chart (from the Australian Climate Atlas, published by the Australian Bureau of Meteorology in 1975). Seasonal rainfall incidence is determined from the ratio (greater/lesser) of the median rainfalls for November to April (austral summer), and May to October (austral winter). The rainfall categories are given in Chapter 4, Table 4.2. Figure 5.1 illustrates the major Australian rainfall zones.

Temperature and humidity

Instead of separate climate zone maps for temperature and humidity, the Bureau of Meteorology provides a combined zonal classification. Table 5.1 gives a description of the six major temperature and humidity zones, and Figure 5.2 is an illustration of these zones.
Figure 5.1 Major Australian rainfall zones (courtesy of the Australian Bureau of Meteorology).
Figure 5.2  Major Australian temperature and humidity zones (courtesy of the Australian Bureau of Meteorology).
Figure 5.3  Major Australian vegetation (Köppen) zones (courtesy of the Australian Bureau of Meteorology).

Climate Classification of Australia

Climate Classes

Equatorial
- rainforest (monsoon)
- savanna

Tropical
- rainforest (persistently wet)
- rainforest (monsoon)
- savanna

Subtropical
- no dry season
- distinctly dry summer
- distinctly dry winter
- moderately dry winter

Desert
- hot (persistently dry)
- hot (summer drought)
- hot (winter drought)
- warm (persistently dry)

Grassland
- hot (persistently dry)
- hot (summer drought)
- hot (winter drought)
- warm (persistently dry)
- warm (summer drought)

Temperate
- no dry season (hot summer)
- moderately dry winter (hot summer)
- distinctly dry (and hot) summer
- no dry season (warm summer)
- moderately dry winter (warm summer)
- distinctly dry (and warm) summer
- no dry season (mild summer)
- distinctly dry (and mild) summer
- no dry season (cool summer)

Based on a modified Köppen classification system. Classification derived from 0.025 x 0.025 degree resolution mean rainfall, mean maximum temperature and mean minimum temperature gridded data. All means are based on a standard 30-year climatology (1961 to 1990). Commonwealth of Australia 2003
Bureau of Meteorology
Table 5.1 Description of the temperature and humidity zone classification.

<table>
<thead>
<tr>
<th>Zone</th>
<th>Zone description</th>
<th>Definition</th>
</tr>
</thead>
</table>
| 1    | hot humid, warm winter     | - average January maximum temp >= 30°C  
                                 - average 3 pm January water vapour pressure >= 2.1 kPa  
                                 (the upper humidity limit for comfort, and where evaporative cooling starts to take effect) |
| 2    | warm humid, mild winter    | - average January maximum temp <= 30°C  
                                 - average 3 pm January water vapour pressure >= 2.1 kPa | |
| 3    | hot dry summer, warm winter| - average January maximum temp > 30°C  
                                 - average 3 pm January water vapour pressure < = 2.1 kPa  
                                 - average July mean temperature > = 14°C |
| 4    | hot dry summer, cold winter| - average January maximum temp > 30°C  
                                 - average 3 pm January water vapour pressure <= 2.1 kPa  
                                 - average July mean temperature < = 14°C |
| 5    | temperate                  | - average January maximum temp <= 30°C  
                                 - average 3 pm January water vapour pressure <= 2.1 kPa  
                                 - average July mean temperature > = 14°C |
| 6    | cool temperate             | - average January maximum temp <= 30°C  
                                 - average 3 pm January water vapour pressure <= 2.1 kPa  
                                 - average July mean temperature < = 14°C |

Köppen vegetation

In the Köppen classification scheme (devised in 1918 by Dr Vladimir Köppen from Austria), the climate of each region is based on temperature and rainfall, as indicated by the native vegetation (Köppen and Geiger 1928). This classification is based on the concept that native vegetation is the best expression of climate, and the climate zone boundaries are selected with vegetation limits in mind (Trewartha 1943). Stern and others (1999) developed the modified Köppen classification that was used for this study. It departs only slightly from the original. The five main Köppen climatic zones found on the Australian mainland and Tasmania are tropical, sub-tropical, desert, grassland, and temperate (the equatorial zone is only just included in the very northern tip of the country). Figure 5.3 illustrates the five main vegetation zones, and the sub-classes within them.
5.2.3.2 Step 2 - Assigning SLAs to climate zone attributes

Each SLA was coded according to its relevant climate zone for the three maps. Although displayed electronically, the climate maps were only available as cartographic elements, not as a vector format. I manually allocated the SLAs to a zone on each map in the following way:

- I printed extra large format maps of the SLA and climate maps at the same scale.
- I overlayed the SLA map onto each climate map and, using a fine pen, traced the attribute boundaries.
- In a GIS program (MapInfo Corporation 1998) I gave each SLA a climate or vegetation zone identity number.

In rural areas, large SLAs sometimes straddled two zones. Following the method adopted by Glass (1995), I determined zone allocation on the basis of (i) land area, and (ii) the position of the largest population centre(s). I assumed that, on average, people in rural SLAs were likely to become infected close to major town centres. Figures 5.4 to 5.6 show the SLAs classified according to each of the three themes (rainfall, temperature/humidity, vegetation).

5.2.3.3 Step 3 - Forming the RRv bioclimatic regions

I grouped SLAs according to their agreement for the three climate and vegetation maps in Excel. This formed 47 homogeneous groups, called “RRv bioclimatic regions” (Figure 5.7). The mean number of SLAs per bioclimatic region was 27 (range 1-291). Regions with the largest numbers of SLAs were located in metropolitan areas. The average area of the RRv bioclimatic regions was 142 667 km² (range 1 - 1 353 568 km²). The RRv bioclimatic regions can be categorised into three size classes:

1. < 50 000 km²
2. 50 - 150 000 km²
3. > 150 000 km².

In the smallest class there were 25 regions. These are all situated within several hundred kilometres of the coast, and have relatively high rainfall. In the second class there were nine regions. In the third class there were thirteen regions. Most of these were situated in remote arid and semi-arid areas.
Figure 5.4  Major Australian rainfall zones, based on Statistical Local Areas.
Figure 5.5 Major Australian temperature and humidity zones, based on Statistical Local Areas.
Figure 5.6  Major Australian vegetation zones (Köppen), based on Statistical Local Areas.
Figure 5.7  Ross River virus bioclimatic regions, based on Statistical Local Areas (n=47).
5.2.4 Qualifications on the method

Several assumptions and limitations underlie the development of the RRv bioclimatic regions:

- I assumed the data were of consistent quality across the continent.
- Macroclimate does not conform to boundaries. It ranges across zones that are stable enough to depict patterns, but are by no means fixed (Bailey 1988). Trewartha wrote that "climatic boundaries, as seen on a map, even when precisely defined, are neither better nor worse than the human judgements that selected them, and the wisdom of those selections is always open to debate" (Trewartha 1943). It follows that the boundaries of the climatic zones (and hence of the RRv bioclimatic regions) will be less reliable than the middle of these areas.
- The boundaries of zones were also the places where SLAs may have been misclassified. Although I took care to work at a high resolution, the method used to assign the SLAs to the climate and vegetation zones may have resulted in some inaccuracies.
- The influence of the three bioclimatic themes (rainfall, temperature/humidity, and vegetation) on mosquito ecosystem boundaries at the level of macroclimate has been well demonstrated, as has their short-term impact on breeding cycles. However, the climate and vegetation zones were defined by criteria that may not necessarily relate to mosquito biology thresholds. It would have been preferable to have had, for example, temperature zones defined in relation to factors favourable or detrimental to supporting mosquito life (such as the number of spring days below zero degrees Celsius).
- The underlying attribute data for each zone of the Köppen system were not available. Some overlap between the Köppen classification and both the other climate zone maps is likely. It is not clear whether this added an unnecessary layer of complication in the selection of the RRv bioclimatic regions. I decided to keep the vegetation classification system in, on the basis that the temperature and humidity zone classes were designed with human thermal comfort foremost in mind. The Köppen classification system relates most directly to the influence of climate on plant growth (and, by extension, biota to sustain vertebrate hosts, and moisture to support mosquito life).
5.3 OUTCOME VARIABLE

5.3.1 Epidemics or incidence?

When developing predictive models for public health purposes, it is important to have a clear understanding of what it is you are attempting to predict. Related studies of associations between climate and mosquito-borne diseases have utilised both epidemics (Gill 1927, Francy and Wagner 1992, Hales et al. 1996, Linthicum et al. 1999, Maelzer et al. 1999) and incidence (Whelan 1989, Tong et al. 1998) as outcome variables in statistical modelling. Table 5.2 summarises those that have used epidemics.

Table 5.2 Definitions of “epidemic” used in related mosquito-borne disease studies.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Study Description</th>
<th>Period</th>
<th>Definition of Epidemic</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Hales et al. 1996)</td>
<td>Global dengue epidemics correlated with average SOI values</td>
<td>fifteen years</td>
<td>A year in which a dengue outbreak was reported in PROMED.</td>
</tr>
<tr>
<td>(Maelzer et al. 1999)</td>
<td>Outbreaks of RRv disease in the south-eastern States of Australia correlated with the SOI</td>
<td>1928 to present</td>
<td>(i) Outbreak reported in the literature, or (ii) more than 600 cases notified (in NSW and Vic.) between the months of January and April.</td>
</tr>
<tr>
<td>(Linthicum et al. 1999)</td>
<td>Analysis of rainfall influence on Rift Valley fever epidemics</td>
<td></td>
<td>A year in which there “RVF viral activity” was recorded.</td>
</tr>
<tr>
<td>(Francy and Wagner 1992)</td>
<td>Developed a predictive model of Western Equine encephalitis epidemics, using weather variables</td>
<td>eleven years</td>
<td>Values more than one standard deviation from the period median.</td>
</tr>
<tr>
<td>(Gill 1927)</td>
<td>The &quot;epidemic figure&quot;, in conjunction with other environmental data, was used to forecast malaria peaks in parts of India</td>
<td>~1910s to 1940s</td>
<td>The number of times the fever mortality in the month of October exceeded the mean monthly fever mortality of the four months from April to July.</td>
</tr>
</tbody>
</table>

The principal argument against the use of “incidence” as the outcome measure for a relatively rare disease such as this relates to their instability in small denominator populations (Hertz-Picciotto 1998, Atkinson and Molesworth 2000), and hence the greater likelihood of extremely high rates based on only a few cases (Hertz-Picciotto 1998). This is a particular problem in sparsely populated non-metropolitan areas, which constituted many of the study regions (Chapters 6 and 7).
A major argument in favour of using "epidemic" as the outcome related to the potential end use of the predictive models (i.e., early warning of RRv disease). Public health resources for the management of RRv disease in non-metropolitan areas are typically severely limited. Given the low health priority assigned to RRv disease in some regions, mosquito control would be employed only when a significant health risk was likely (Terence Carvan pers. comm. 2000). Thus, advance warning of the probability of a very large (i.e., epidemic) RRv disease season, rather than an increase in cases from some nominal amount to another, better reflects the level of resources available for administration of an early warning system, surveillance and response, and seems intuitively easier to apply in a policy context.

A second reason to use "epidemic" as the outcome related to the NNDSS data quality. Under-reporting during RRv disease outbreaks is considered common (see Chapter 3). There was no way of assessing whether there was differential under-reporting between SLAs. In an extreme situation, this could result in under-estimation of the true number of epidemics in an SLA over the period. It is more likely that the occurrence of an epidemic number of cases over a season would be reflected with some accuracy. As noted in Chapter 4, there is some evidence for differential reporting periods (i.e., time between onset of symptoms and report) between States. In States with a longer period between disease onset and report, the estimated date of infection, and hence of the temporal sequence of infections, may be less precise (not such a problem if "epidemic" is used as the outcome).

In summary, both epidemics and incidence rates are valid outcome measures for the predictive modelling of mosquito-borne diseases. Given the aims of this study, and to minimise the data quality issues, "epidemic" was favoured as the outcome variable.

5.3.2 Definition of an epidemic

Last (2001) defines an epidemic as "the occurrence in a community or a region of cases of disease clearly in excess of normal expectancy". The American Public Health Association (Chin 2000) has expanded this definition to include the notions of spatial boundary, time, and affected populations. The number of cases that constitutes an epidemic can vary, and will depend on the agent of infection, the size and type of population exposed, previous experience or lack of exposure to the disease, and the time and place of occurrence.
Some method of assessing the background level of disease is needed to determine whether a cluster of cases is in excess of normal. The World Health Organization defines an outbreak as being when the “observed number of cases exceeds the expected number of cases in a given population for a given period” (Grein et al. 2000). From a statistical point of view, the less sensitive a definition (i.e., only very high numbers of cases are considered an epidemic) the more information that is lost about the relationship between the exposure and the outcome. Conversely, a highly sensitive definition (where fewer cases constitute an epidemic) makes it more difficult to determine how much the exposure variables contribute to changes in the outcome.

Francy and Wagner (1992) developed a model to predict outbreaks of Western Equine encephalitis. Of related studies, their definition of “epidemic” seems most appropriate for the RRv disease analyses: “values which are more than one standard deviation from the period median”. The mean or median are common measures of central tendency, and the standard deviation of the spread of a distribution. The mean estimates the average number of events that have occurred, and is sensitive to extreme values, which distort representation of the typical value of the data. The median, another possible measure of the middle of the data, is not affected by extreme values. However, the median value is determined solely by rank, and does not provide information about the other values in a distribution (such as whether they are unusually high or not). On balance, the mean has the advantage of being a more frequently employed measure of variability (Hennekens and Buring 1987). Given the possibility of under-reporting in epidemic years, a conservative cut off - one standard deviation above the mean rather than two standard deviations - was chosen.

For the purposes of predictive modelling, an epidemic was defined as:

the number of cases in any one RRv disease year (July to June) that exceeded the mean plus one standard deviation of all cases recorded in an SLA during the study period (July 1991-June 1999).

The standard deviation is based on the assumption that incidence has a Poisson distribution. The definition of the outcome variable (dichotomous) was:

whether or not an epidemic occurred in an SLA in one year.
5.4 Predictive modelling

5.4.1 Introduction

As the geographical classification identified 47 distinct groups, which by definition represent different RRv bioclimatic regions, theoretically it was possible to consider 47 models. Practically, it was necessary to identify two of these regions to use as case studies for modelling purposes. With the first study area, the aim was to determine whether prediction using climate data was possible. With the second, the aim was to test whether the methods developed in the first study area could be replicated in another.

Myers and others (2000) identify two basic methods for modelling the mosquito-borne disease and environment relationship: (i) biological (also called “process-based”: McMichael et al. 2000), and (ii) statistical. Both employ fundamentally different approaches. The biological approach incorporates a literature-derived equation about vectorial capacity and its relationship to meteorological variables. It requires “details on all parameters and variables considered to be important in transmission” (Myers et al. 2000). These models can, theoretically, be used in situations where the external factors change, as long as the impact of these changes on the transmission parameters is established. For example, if the extrinsic incubation period for a virus is determined to be 5 days at 25°C, 4 days at 28°C, 3 days at 32°C, etc, then the effect of increasing temperatures due to climate change on disease infection rates can be estimated. The problem with this approach, for Ross River virus disease and with many of the lesser studied mosquito-borne diseases, is that the transmission parameters for most of the vector species are poorly understood. Even for the three major RRv vector species there are substantial gaps in our understanding of the effect of temperature on extrinsic incubation periods and breeding times (Woodruff et al. 2001).

In the absence of such information, statistical modelling is a second option. This approach uses information about environmental exposures to establish a relationship with epidemics of disease (or case numbers). In terms of mosquito-borne diseases, the statistical model derives an empirical statistical equation from the current distribution of a disease in terms of its known response to climatic determinants (such as temperature and rainfall) (Myers et al. 2000). This approach assumes that this relationship will remain static, which is a problem in terms of climate change. Current models of the environment-
disease relationship may be less reliable in future, if projected changes to temperature and rainfall occur (McMichael et al. 2000).

Myers and others (2000) note the importance of predictive models incorporating both the extrinsic and intrinsic factors that drive mosquito-borne diseases. Intrinsic factors include vertebrate host susceptibility, infection, etc., that change over time in a normal population. Extrinsic factors include climate (which affects the average amount of transmission in a region) and weather (which influences the occurrence of disease within seasons). Statistical models are designed to incorporate the impact of extrinsic factors, but do not automatically cope well with the intrinsic factors.

This thesis uses a statistical approach to develop predictive models for RRv disease in each of the study regions. In addition to including the relevant extrinsic factors that affect transmission of RRv (in this case, climate and environment), the models also attempt to account for important intrinsic factors that can interrupt the occurrence of epidemics (i.e., vertebrate host immunity levels).

5.4.2 Criteria for selecting the study areas

This section presents the method for selecting which RRv bioclimatic regions would be the focus of final investigation. I established a set of criteria to help select the most appropriate two case study areas. These included "important" and "desirable" attributes.

5.4.2.1 Important criteria

1. An area should comprise a large enough sample (i.e., number of SLAs) to enable a real association between climate and epidemic occurrence to be detected.
   
   Because "epidemics" are based on SLAs, RRv bioclimatic regions with only a few SLAs would be very hard to develop reliable models for.

2. An historical record of RRv disease outbreaks.

   In addition to the descriptive epidemiology of RRv disease during the study period, an historical record of disease events provided more extensive information about the season of onset, inter-annual fluctuations between outbreaks, anomalous outbreaks, observed weather events prior to the onset of cases, and coincident vector and vertebrate host species in an area. For the
purposes of hypothesis development, the amount of information about the epidemiology of the virus in an area was a persuasive argument for selection.

3 A record of RRv vector studies in an area.
   No geographical map of mosquito species distribution exists for Australia. The knowledge base of RRv vectors for this study came from an extensive review of the literature, and discussions with entomologists. Regional information on species (particularly if recent) was essential for understanding transmission. To generate testable hypotheses for an area, the following were needed: (i) knowledge of the major vector species and the extent of their role in RRv transmission (i.e., does a species amplify virus within the vertebrate population, transmit virus to humans, or do both?), and (ii) field or laboratory studies of the effect of temperature, humidity, and rainfall on the breeding and survival of regional RRv species.

4 Contact with mosquito experts in an area.
   Local mosquito expertise was important for filling the gaps in the literature, and for "ground-truthing" model findings. I developed contacts with many local experts in Australia (including entomologists, and mosquito control managers working for Local Councils and State Health Departments). For a number of reasons, these contacts were more productive in some areas than others.

5 Routine mosquito surveillance data.
   This study did not initially aim to examine the contribution of mosquito surveillance data as a predictor of RRv epidemics. Findings from the Murray study area (Chapter 6) indicated that mosquito density should be tested as a model variable in a second study region. This became an essential criterion for the selection of that region.

6 Variance in case distribution between the years of the study.
   In some RRv disease endemic regions (such as in parts of Queensland), high annual case numbers were reported in each year of the study, without enough variation between years to identify epidemic and non-epidemic periods. Epidemic regions (where the inter-annual cycle is more prominent) were more amenable to statistical analysis than endemic ones.
5.4.2.2 Desirable criteria

1. Was the area strongly affected by the ENSO phenomenon?
   The influence of ENSO on rainfall patterns varies substantially across Australia. To test its potential value as a predictor variable in an early warning system, areas of the country that were strongly affected by ENSO were preferred to those that were not affected.

2. Information on the vertebrate hosts.
   Vertebrate host immunity is an important moderator of RRv transmission to humans. Studies of potential RRv hosts have been very few. These have aimed to classify vertebrates as effective reservoir hosts for the virus. No studies have examined the life cycle of primary reservoir hosts in relation to population turnover and RRv immunity levels. General information in a region pertaining to the life cycle of vertebrate hosts was the best that could be hoped for.

From these criteria, six geographic areas of Australia were suggested as potential study sites. Table 5.3 summarises the merits of each in relation to the criteria. From this group, the Murray River area of south-eastern Australia, and the south-west of Western Australia best satisfied the criteria.

Table 5.3 Summary of potential RRv disease study areas against the selection criteria.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Potential study areas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Murray River</td>
</tr>
<tr>
<td>Important</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>✓</td>
</tr>
<tr>
<td>Record of disease outbreaks</td>
<td>✓</td>
</tr>
<tr>
<td>RRv vector studies</td>
<td>✓</td>
</tr>
<tr>
<td>Contact with mosquito experts</td>
<td>✓</td>
</tr>
<tr>
<td>Mosquito trapping data</td>
<td>X</td>
</tr>
<tr>
<td>Epidemic variation</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5</strong></td>
</tr>
<tr>
<td>Desirable</td>
<td></td>
</tr>
<tr>
<td>Strongly ENSO affected</td>
<td>✓</td>
</tr>
<tr>
<td>RRv vertebrate host information</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2</strong></td>
</tr>
</tbody>
</table>

? = Unknown
5.4.3 The two case study areas

5.4.3.1 Murray River study area

- The Murray River forms part of the largest south-eastern river system. This area has the longest and largest documented record of RRv disease outbreaks in Australia.
- The Murray River geographical area contains two RRv bioclimatic regions. Statistical modelling was conducted in both these regions, because the area is intrinsically geographically linked. Henceforth, they are referred to as “Region 1” or “Region 2” in the “Murray study area”.
- Given the extensive history of disease outbreaks, a comparatively large amount of observational research on RRv transmission (vectors, seasonal cycle, etc.) has been conducted in these regions.
- There was high variability in the disease pattern in these regions during the study period (epidemics predominantly occurred in two of the eight years in many SLAs, with few epidemics in intervening years).
- At the commencement of the thesis research, I made contact with staff from the Greater Murray Health Service (responsible for disease control for the New South Wales side of the Murray River), and they expressed a strong interest in supporting the study. Terence Carvan helpfully provided advice on RRv in the region.
- Rainfall patterns in this part of Australia (i.e., the south-eastern corner) are strongly affected by ENSO. There was a very small amount of information on the types of vertebrate hosts in the Murray River area.

5.4.3.2 Southwest study area

- The south-west of Western Australia contained one bioclimatic region (henceforth referred to as the “Southwest study area”).
- Large outbreaks of RRv disease have been reported in the Southwest from time to time since the mid-1980s, and as a result the ecology of the virus has been substantially and consistently documented since then (e.g. vector trapping and disease surveillance was conducted in most years of the study period). Dr Michael Lindsay (pers. comm. 2002) from the University of Western Australia generously provided information on the RRv transmission cycle, and helped in the assemblage of mosquito surveillance data.
- There was high variability in the epidemic pattern in this area (epidemics occurred in several of the years, with few epidemics in intervening years).
Chapter 5 – Case study methods

- Historical meteorological observations indicate that this area is affected by the ENSO phenomenon, although not as strongly as the south-eastern part of the country. Several RRv observational studies have noted an association between higher than average mean sea levels and high numbers of mosquitoes or case incidence in the Southwest (Lindsay 1995, Lindsay et al. 1996c).

The two disease modelling chapters (6 and 7) provide a descriptive profile of these study areas.

5.4.4 SLA exclusion criteria

SLAs with either of the following characteristics were excluded from the statistical modelling analysis:
- Located in a metropolitan region.
- Ten or fewer notifications of RRv disease during the study period.

As discussed in Chapter 3, to compensate for the limitations of the notification data (which do not record place of infection), I assumed that in rural areas with geographically large SLAs, infection occurred in the same SLA as the SLA of residence. This assumption cannot be true for large cities, where the majority of RRv notifications are imported. Thus, metropolitan SLAs within the study areas were removed prior to analysis. The boundaries for the metropolitan regions were obtained from the Australian Bureau of Statistics (1998a).

I also excluded any SLA that reported ten or fewer RRv disease notifications during the study period. My reasons were threefold. First, a precondition of RRv transmission is that both vectors and virus must be present within an area. If an SLA did not record any notifications during the period, that may have been because local weather conditions were not suitable to sustain RRv vectors. It may also have been because there were not any RRv vectors and virus in the SLA. Given the paucity of RRv vector and viral records for the country as a whole, I was not able to assess whether viral activity or RRv vector species were present in all SLAs. Second, the assumption that infection and place of residence was the same for most SLAs is likely to be generally true. However, when only a very small number of notifications were reported in a rural SLA, there was a reasonable chance that these may have been imported. Third, it was methodologically problematic to
develop a definition of an epidemic using very low numbers spread over eight years (e.g. 1, 0, 0, 0, 3, 0, 2, 0). Note this exclusion only applied to SLAs for the predictive modelling analyses. The national analysis of RRv disease (Chapter 4) reported on all notified cases in Australia (i.e., SLAs n=1272). It is possible that some RRv activity in the excluded SLAs (occurring at very low levels for whatever reason) may have occurred during, or prior to, the study period. A limitation of this approach was the potential loss of information about the climate-disease relationship in such SLAs.

5.4.5 The modelling method

Two predictive models were developed for each study area. The “early warning” models comprised variables for the months July to November (late austral winter to the end of spring). Initial cases of RRv disease are typically identified around November in both the Murray and the Southwest study areas. Early warning of the likelihood of an epidemic season in November would be timely from the point of view of mosquito control and prevention education.

In the Murray study area, the “late warning” models included variables for December to February (austral summer) in addition to those from July to November. Although substantial cases have usually been notified by February, in most years the numbers of cases peak between March and April. As the interval between infection and symptoms is often as low as seven to nine days (Fraser and Cunningham 1980), it would still be useful to check the probability of an epidemic at this time, and to issue more public alerts if the results were conclusive. In the Southwest study area, cases peak on average one month earlier, and variables for July to January were used.

Prior to developing the models, a literature review was conducted of the vector and host ecology, and the epidemiology of RRv disease was described for each study area. This information formed the basis of the hypotheses that were developed (described in Chapters 6 and 7).
Chapter 5 - Case study methods

The main steps taken to develop the predictive models in each study area (discussed more fully below) were to:

1. Conduct a single variable regression analysis, and identify candidate variables.
2. Develop an early warning multivariable logistic regression model, using candidate variables for the months 1-5.
3. Develop a late warning multivariable logistic regression model, using candidate variables for the months 1-8 in the Murray area, and 1-7 in the Southwest area.
4. Choose best-fitting models (according to established criteria).
5. Cross-validate the models within the data set.

Analyses were undertaken using the statistical package Stata 6.0. (StataCorp. 1999). Table 5.4 summarises the definitional terms relating to the study design.

Table 5.4  Definitional terms relating to the study design for this thesis.

<table>
<thead>
<tr>
<th>Study terms</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study period</td>
<td>01 July 1991 to 30 June 1999.</td>
</tr>
<tr>
<td>One year</td>
<td>One 'RRv disease' year - July (month 1) to June (month 12)</td>
</tr>
<tr>
<td>Spatial unit</td>
<td>Statistical Local Area (SLA)</td>
</tr>
<tr>
<td>Temporal unit</td>
<td>One calendar month</td>
</tr>
<tr>
<td>Outcome</td>
<td>Whether or not an epidemic occurred in an SLA in a year</td>
</tr>
<tr>
<td>Epidemic</td>
<td>The number of cases in any one year that exceeded the mean plus one standard deviation of all cases recorded in an SLA during the study period</td>
</tr>
</tbody>
</table>

5.4.5.1  Step 1 - Single variable analysis

Chapter 3 provides full details of the climate and environmental data used in this study. Table 5.5 is a list of the explanatory variables considered in the single variable modelling. For each type of variable (e.g. total rainfall) there was a separate variable for the months of July to February in the Murray area, and July to January in the Southwest area (reflecting the different timing of epidemics between the areas).
Table 5.5 Climate and environmental variables tested in the predictive models.

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Unit</th>
<th>Variable (July, August, Sept...Jan/Feb)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rainfall</td>
<td>mm</td>
<td>• Total rainfall</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Number of rain days</td>
</tr>
<tr>
<td>Temperature</td>
<td>°C</td>
<td>• Average temperature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mean maximum temperature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Absolute maximum temperature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mean minimum temperature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Absolute minimum temperature</td>
</tr>
<tr>
<td>Evaporation</td>
<td>mm</td>
<td>• Evaporation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Maximum evaporation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Minimum evaporation</td>
</tr>
<tr>
<td>Humidity</td>
<td>%</td>
<td>• Average relative humidity at mean temperature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Average relative humidity at maximum temperature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Average relative humidity at minimum temperature</td>
</tr>
<tr>
<td>Vapour pressure</td>
<td>hPa</td>
<td>• Vapour pressure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Maximum vapour pressure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Minimum vapour pressure</td>
</tr>
<tr>
<td>SOI</td>
<td>–</td>
<td>• Southern Oscillation Index</td>
</tr>
<tr>
<td>SST</td>
<td>°C</td>
<td>• Sea surface temperature</td>
</tr>
<tr>
<td>Tide</td>
<td>cm</td>
<td>• Absolute tide height (Southwest area)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Average tide height (Southwest area)</td>
</tr>
<tr>
<td>Mosquito</td>
<td>–</td>
<td>• Mean mosquito density (Southwest area**)</td>
</tr>
</tbody>
</table>

* Depending on the study area, there were 7 or 8 variables for each exposure group. For example, "total rainfall" comprised "total rainfall July", total rainfall August", ... , "total rainfall February".

** Mosquito data were only available for two months (November and December).

The variables were regressed one at a time against the outcome (occurrence of an epidemic), using the logistic command in Stata. Variables with a P< 0.2 were classified as "candidate variables", and were considered in the multivariable logistic regression modelling.

Given the interrelationship between climate variables, I computed a correlation matrix for all variables that were significant in single variable analysis. Stata software manages highly correlated variables in the following manner: one variable remains statistically significant when the contribution of both is modelled, and the other (which is no longer significant) is automatically dropped. Thus, highly correlated variables were noted, but were not removed from the list of potential candidate variables.
In an attempt to better describe the climate-disease relationship, several variables were derived from those listed in Table 5.5. Averages were derived of two or more variables if (i) they were consecutive in temporal sequence (e.g. rainfall for August, September, and October), and (ii) if each was significant in the single variable analysis. Lags of one year were derived for the rainfall variables, to test hypotheses relating to vertebrate host immunity.

Logistic regression constrains the covariates to follow a logistic pattern (Greenland 1998). Since not all covariates may follow that pattern, I explored non-linearity in the relationship between epidemic occurrence and the major predictor variables in two ways:

- by grouping the values of continuous variables into categorical ones, to ascertain whether ‘threshold’ values operated, and
- by adding polynomial terms (i.e., $x + x^2, x + x^3$, etc., where $x$ is a covariate).

A detailed explanation of the variables derived for analysis within each of the study areas is covered in the relevant chapters (6 and 7).

5.4.5.2 Steps 2 and 3 – Early and late warning models

This study had a spatio-temporal dataset, with time components (epidemics by year) and spatial components (epidemics by SLA) in the matrix. The central hypothesis of the models was that epidemics can be predicted by some function of climate and of time (month) in each study area.

Two models were developed for each of the study areas. The early warning model used all candidate variables for the months 1-5 (July to November). The late warning model used all candidate variables for the months 1-8 (July to February) in the Murray study areas, and months 1-7 (July to January) in the Southwest study area.

Each of the variables was adjusted for all of the others by performing multivariable logistic regression, using the Logistic option in Stata. The probability of an epidemic year was expressed in terms of the climate variables by a logistic regression model. The equation for the model had the form:

$$\log \left[ \frac{P}{(1-P)} \right] = a + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

where $P$ is the probability of an epidemic, the $x_i$ are explanatory variables specific to an SLA, and $a, \beta_1, \ldots, \beta_n$ are constants that are estimated from the data.
This study had a rich explanatory data set. Given the large number of variables (18-21, depending on the study area) and number of months for each variable (7 or 8), there was a large combination of factors that could be reviewed for a predictive association. However, there were constraints on the number of predictor variables that could be entered into Stata at one time. The purpose was to develop models with biologically plausible variables that had a significant relationship with the outcome. I took a pragmatic approach to this analysis, and entered the candidate variables (i.e., those remaining after single variable analysis screening) in biologically and chronologically relevant groups. For example, spring variables would be entered before summer variables. When a working model was developed, variables were substituted, one by one, with similar variables to determine which was the better predictor (for example, “total rainfall” in a month was replaced with “rain days” in the same month). To determine the appropriate combination of variables to test in the modelling, I was guided by (i) the evidence from the literature (e.g. temperature thresholds for breeding), (ii) the hypotheses that I developed for each study area relating to the influence of climate on vector and host dynamics, (iii) the results of the single variable analyses, and (iv) my own judgement.

Using Stata’s automatic backward stepwise procedure, variables that were highly correlated or that had a P-value of more than 0.2 were removed. The model was then re-fitted using those variables that remained statistically significant, and another group of variables would be included. Variables that were essential to the transmission cycle (discussed in Chapter 1) were retained in the model even if they were non-significant.

5.4.5.3 \textit{Step 4 – Criteria for best-fitting models}

I assessed the models to determine how effective they were in describing the probability of an epidemic. Preliminary criteria for judging a satisfactory model were:

(i) More than 90% accuracy between the observed and fitted values for the unvalidated model. I considered a value of the predicted probability of an epidemic (which lay somewhere between 0 and 1) to be an accurate reflection of an epidemic if it was greater than 0.75, and an accurate reflection of a non-epidemic if it was below 0.25.

(ii) Significance on a Hosmer-Lemeshow test, a goodness-of-fit test available for binary response models. This test groups the observed and expected number of responses based on deciles of the estimated probabilities (Hosmer and Lemeshow 1980). If the
P-value is large, then the model fits the data well. This test is preferred to the Pearson test when the number of the observations per covariate pattern is small (Hosmer and Lemeshow 1980).

(iii) An additional criterion was the degree of spatial correlation of the model residuals. Stata provides diagnostic techniques that I used to check the spread of the residuals.

Given the purpose of this study, these criteria alone were not sufficient to indicate a successful prediction model. Ultimately, the crucial test was the results of the validation test (discussed below).

5.4.5.4 Step 5 - Cross-validation

I validated the models' predictive ability using a method of cross-validation. This method used the same group of SLAs and the same time-period to validate the predictive ability of the model (following Chey et al. 1999). The whole sample was divided into the eight study years. In a rotating fashion, one year was systematically dropped and the remaining seven years were used to derive the parameter estimates for the model. I used these parameter estimates to calculate the predicted probabilities of epidemics in the remaining one-eighth of the data. I repeated this procedure until the predicted probability of epidemics for all the years was calculated.

To examine the validity of a model's predictions, I calculated the accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the model at predicting epidemics for each year during the study, and for the period overall. Hereafter, this method is referred to as the "validation test". The logistic regression model output for the probability of an epidemic was a continuous variable, and thus the sensitivity and specificity could be varied by changing the cut-off value that differentiated positive from negative outcomes. I calculated the accuracy of the model output for all years (i.e. the percentage of epidemics and non-epidemics that were accurately predicted) for cut-off values of 0.5, 0.7, and 0.9. This showed how severe the

---

4 Validation Test. Sensitivity: the proportion of epidemics that the model correctly identified as epidemics; specificity: the proportion of non-epidemics that the model correctly identified as non-epidemics; positive predictive value: the likelihood that an epidemic predicted by the model really is an epidemic; negative predictive value: the likelihood that a non-epidemic predicted really is a non-epidemic. Accuracy: the percentage of epidemics and non-epidemics correctly predicted. Except where otherwise noted, a cut-off probability of 0.5 was used to define an epidemic.
trade-off was between sensitivity and specificity, and was used to discuss the benefits of varying thresholds for defining the probability of an epidemic.

In view of the fact that the validation method required dropping one year for each validation cycle, there were some cases where epidemics and non-epidemics were predicted completely. This was expected, given the limited data series relative to the number of variables.

5.4.5.5  Combined region analysis in the Murray study area

In the Murray study area only, the data for the two regions were pooled and a “combined model” was built for the early warning and late warning periods. The purpose of the combined model was to examine whether the RRv bioclimatic regions were a useful method for grouping SLAs into climatically homogeneous groupings. If a prediction model could be developed for the combined regions that was superior to the two individual models, this would have suggested that the bioclimatic system was not a very useful method for clustering SLAs (i.e., choosing the boundaries of spatial analysis).

The questions that guided the combined regions analysis were:

1  Could epidemic prediction with a similar or better performance (i.e., accuracy and reliability) be achieved at a larger spatial scale? Could a better coefficient estimate be obtained for variables that operated in essentially the same manner in both regions?
2  Did the combined model increase parsimony (compared to the separate region models)? Were there any variables that contributed little to the model that could be removed?

The method used was to conduct a backwards stepwise regression with the variables from the final (early or late) models for each region. Variables that were dropped in this process were recorded. Variables with significantly different coefficients were given a separate coefficient to denote “region” (implemented by an interaction term in Stata). “Global” variables were identified (i.e., those variables shared by both regions). The Log Likelihood test was used to determine whether these global variables operated in a significantly different way in the two regions. If this was the case, separate coefficients were estimated for the variables for each region. If the variable showed no regional effect, data for the variable were pooled and a single coefficient was estimated.
5.4.5.6 Managing dependency in the data

Pfeiffer (1996) has classified two types of spatial effects relevant to the modelling of spatial phenomena. First order effects are large scale trends, where there is overall variation in the mean value of a parameter. The development of the RRv bioclimatic regions was premised on an assumption that regional synchrony in long-term average climate factors correlates with synchrony in mosquito habitat – thus influencing species type, transmission cycles, etc.

Second order effects are produced by spatial dependence, which needs to be managed to provide an approximation of existing effects. These effects are local, and represent the tendency of neighbouring values to follow each other in terms of their deviation from the mean (Pfeiffer 1996). Co-located SLAs can be expected to have significant agreement in, for example, rainfall amount and pattern. This agreement would usually be expected to decline with increasing distance between SLAs (Koenig 1999).

First order effects can be modelled using standard regression techniques. Second order effects violate the assumption of standard statistical analysis, which assumes independence between covariates. To account for the covariance structure in the data giving arise to these local effects, I introduced other variables into the models. These are detailed in the relevant chapters (6 and 7). To check for additional dependence, over and above that which was removed by these variables, I also checked the covariance structure of the model residuals. Any correlation in the residuals was likely to be unbiased, but may affect the reliability of the width of the confidence intervals.

5.4.6 Confounding and interaction

The periodic nature of epidemics can be due to many non-climatic factors, which need to be identified and managed in analysis to ensure that the association between event and disease is not artefactually enhanced or reduced.

5.4.6.1 Mosquito control activities

Changes in public health management of RRv disease over time, such as mosquito control and campaigns aimed at increasing public practice of personal protection activities, may reduce human population vulnerability. Unless adjusted for in analysis, areas where mosquito control is routinely and effectively practised may reduce total
mosquito population numbers and hence the likelihood of epidemics in otherwise favourable years. To summarise from Chapter 1, the effectiveness of mosquito control can be severely affected by the timing of spraying, the need for detailed and updated information on local RRv species. Even in the very population dense regions where studies have been conducted, there was uncertainty about the effectiveness of particular spraying programs.

In the Murray study area, no mosquito control activities were conducted on the northern border of the Murray River (part of Region 1). Limited interventions (i.e., sporadic larviciding) were conducted in some of the southern SLAs. In the Southwest study area, low-level adulticiding was conducted in some of the northern SLAs where mosquito trapping occurred. In light of the evidence presented in Chapter 1, and given the low level of mosquito control and large geographic areas over which spraying was conducted, it seems reasonable to assume that the affect of vector control on the outcome was negligible.

5.4.6.2 Irrigated regions

Undrained irrigated areas may create ephemeral sites for mosquito breeding, and in years when climatic conditions are suitable for an epidemic they could contribute to the number of available breeding sites. The predominant irrigation method used by cropping industries along the Murray River is water releases from the Hume Reservoir throughout the growing season (typically September-April). The Murray-Darling Basin Commission provided data on flow from the Reservoir (mega litres/day) that I used to interpret the models in the Murray study area. There was no information to suggest that irrigated crops influence mosquito breeding cycles in the Southwest area (Lindsay 1995).

5.4.6.3 Human immunity

Humans, once infected, are usually immune for life. As expected, the sero-prevalence studies conducted to date show an increase in population immunity levels by age group (for a review of these, see Harley et al. 2001). The results of these studies vary substantially, however, and cannot be generalised to other regions. For the purposes of this study, and in recognition of the fact that non-resident cases infected in a region (such as tourists) would not be recorded in that region, I assumed that the level of human population immunity would not change appreciably from year to year. This was based on a second assumption – that the number of residents in each unit area was large and
that population turnover occurred. The population size of the base unit (an SLA) was, on average, large (mean 23,700, range 2,106 to 99,098). The Social Atlas of Rural and Regional Australia (Haberkorn et al. 1999) records changes in the Murray study area populations during the study period of between -5% and +5%. In the Southwest study area, there was a 5-20% population increase.
Chapter 6

The Murray River study area: using climate data to predict epidemics

6.1 Abstract

This chapter describes the influence of climate on epidemics of RRv disease for two RRv bioclimatic regions that surround the Murray River of south-eastern Australia. The objective was to test if climatic factors alone could be used to provide timely prediction of the probability of an epidemic, with a sufficient degree of accuracy (more than 70%). In each region, two predictive models were developed: an “early warning” model (using explanatory variables for the months of July through to November) and a “late warning” model (variables for the months of July through to February). Two patterns were evident: epidemics occurred following high summer rainfalls or high winter rainfalls. Low late spring temperatures, and high late summer temperatures were also necessary. A pre-condition, relating to host-virus dynamics, was lower than average spring rainfall in the pre-epidemic year. The greatest model sensitivity for Region 1 was 96% and for Region 2 was 73%.

I conclude that early warning of climate conditions conducive to outbreaks of RRv disease is possible at a regional level with a high degree of accuracy, and with sufficient time for public health response. While climate is pivotal in generating the conditions needed for epidemics of RRv disease, human influence (through environmental modification) can also duplicate these conditions. These models may have application as
a decision tool for health authorities to use in risk management planning. The results presented in this chapter were published in Epidemiology (Woodruff et al. 2002).

6.2 PROFILE OF THE STUDY AREA

6.2.1 Population and land use

The rationale for choosing the Murray River area for study is outlined in Chapter 5. This geographical area (Figure 6.1) contains two RRv bioclimatic regions: “Region 1” and “Region 2”. These two regions extend from Adelaide in the west to Gippsland in the east, and surround the Murray River, the largest water source in south-eastern Australia. The River is bordered by New South Wales and Victoria, and flows through South Australia.

Following the method outlined in Chapter 5, a number of Statistical Local Areas (SLAs) in each of the two RRv bioclimatic regions were excluded from analysis:

(i) Six SLAs attached to the Adelaide metropolitan Statistical Divisions (as defined by the ABS) in Region 2 were excluded,
(ii) 196 SLAs that recorded less than 10 cases of RRv disease during the period were excluded from Region 1, and 29 from Region 2, and
(iii) Seven SLAs that were geographically isolated in Region 2 were removed (from Tasmania).

Figure 6.2 shows the two bioclimatic zones and the study regions contained within them. Figure 6.3 is a detailed map of the two regions, with major towns shown.

Table 6.1 Population characteristics of the two Murray study regions.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Region 1</th>
<th>Region 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of SLAs</td>
<td>14</td>
<td>24</td>
</tr>
<tr>
<td>Population*</td>
<td>111 924</td>
<td>788 679</td>
</tr>
<tr>
<td>Area (km²)</td>
<td>84 358</td>
<td>40 808</td>
</tr>
<tr>
<td>Population density (per km²)</td>
<td>1.2</td>
<td>19.3</td>
</tr>
</tbody>
</table>

Region 1 comprised 14 SLAs (total area 84 358 km²) and Region 2 comprised 24 (total area 40 808 km²) (Table 6.1). In 1996, the total population of Region 1 was 104 017, and of Region 2 was 788 679. The populations of the SLAs in the two regions were quite different. In Region 1 (located generally to the north of the Murray River, with a drier
Figure 6.1  Australia, showing the location of the Murray study area (shaded).
Figure 6.2  The two RRv bioclimatic regions of the Murray River (light shading), and the two Murray study regions contained within them (dark shading).
Figure 6.3  The Murray River study area: Regions 1 and 2. Major towns are highlighted.
climate and less productive land), the average population size per SLA was 7995 (range 455 - 43 579). In Region 2 (higher rainfall and more fertile land), the average population size per SLA was 32 862 (range 3 983 - 117 225).

Principal land use in both regions is mainly rural, with some cities and large towns. In those SLAs that border the Murray River, irrigation is a dominant feature for rice and horticultural production. The rest of the regions are predominantly used for cropping and grazing. Some irrigation practices, in particular the annual releases of water from the Hume Reservoir at the upper reaches of the Murray River, could at times provide impermanent breeding sites for mosquitoes. The amount of water drawn from the Murray Basin for irrigation is nearly at the limit of the water resource, however, and the resulting environmental impacts (soil loss, salinity, acidification, and water pollution through pesticide and herbicide usage) are causing a major re-assessment of land use practices (State of the Environment Advisory Council 1996b). In general, the improved irrigation practices in this region have dramatically reduced the availability of water sites for mosquito breeding in the Murray area (Dhileepan 1996).

6.2.2 Climate and environment

Region 1 is on the edge of the large internal landmass of Australia, and is generally categorised as semi-arid. It has hot dry summers and cold winters. Region 2 (further south and closer to the moderating influence of the coast) is temperate. The vegetation of Region 1 is characterised by grasslands, and of Region 2 by temperate forests and shrubs.

Seasonal maximum and minimum temperatures vary most in Region 1 due to the absence of the stabilising influence of the oceans on temperature fluctuations (see Figure 6.4 for average temperature profiles). Summers are warm to very hot in Region 1 (mean minimum temperatures are 15°C across the region, and mean maximum temperatures are 31°C). Winters are cold (mean minimum temperatures are 4°C, and mean maximum temperatures are 15°C). In Region 2, temperatures are slightly less extreme: summers are warm (mean minimum temperatures are 13°C, and mean maximum temperatures are 27°C). Winters are cold (mean minimum temperatures are 5°C, and mean maximum temperatures are 14°C). In both regions, daily maximum temperatures can exceed 40°C in summer.
Average yearly rainfall is much lower in Region 1 (340mm) than Region 2 (560mm). The majority of rainfall occurs in winter and early spring in both regions, with only 25% falling during the hottest months (December-March) in Region 1, and 20% in Region 2. Rainfall varies substantially from year to year, however, and heavy falls have been recorded in summer, whilst long drought periods can occur. Very high evaporation in the summer months makes the timing and pattern of rainfalls in these regions integral to mosquito survival.

Rainfall over the Murray River area of Australia has been moderately to strongly related to an El Niño (Chapter 1, Figure 1.2), with an estimated 20-30% decrease in rain overall during such an event (Partridge 1994). La Niña events also strongly influence rainfalls in the region, with an estimated average increase of 30-40% above normal in rainfalls in a La Niña year.

### 6.3 Ecology of RRV in the Murray

#### 6.3.1 Vectors

Russell (1994) considered *Ae. vigilax* to be the principal initiating RRV vector in the Murray River area. *Ae. sagax* and *Ae. vittiger* have also been implicated as minor initiating vectors (Russell 1994). Little definitive research has been conducted to link ambient temperature and rainfall to *Ae. vigilax* breeding, although “clearly the extensiveness of both freshwater and salt marsh larval habitats is linked to rainfall” (Kay and Jennings 2002). It is hypothesised that *Aedes* lay their eggs in river mud, and that following heavy winter or early spring rain, river levels rise and inundate the breeding sites. These infected mosquitoes may then, in turn, infect vertebrate hosts, thus initiating the virus amplification cycle in the Murray Valley area.

Although aedine species are generally restricted to salt-water coastal areas, both *Ae. camptorynchus* and *Ae. vigilax* have been observed to inhabit the saline areas of the lower Murray around Mildura (Lee et al. 1989) and to breed in salt pans that receive irrigation run-off (Dobrotworsky 1965, Kay and Aaskov 1989). All stages of the *Aedes* lifecycle have been collected during cold winters in Victoria (Kay and Aaskov 1989). Marshall and Miles (1984) have suggested that transovarial transmission of virus in *Aedes* species is an overwintering mechanism in eastern Australia.
Figure 6.4 Long-term average monthly rainfall and maximum and minimum temperature profiles for the Murray study regions. Average maximum temperature (°C) ▲; average minimum temperature (°C) ■; rainfall (mm) X.

Region 1

Month

Region 2

Month
Dhileepan (1996) assessed the species composition and abundance of adult mosquitoes in the Murray Valley area during 1991 to 1995 (the first four years of this study) in relation to the altered pattern of irrigation in the area (see Table 6.2 for the results of that trapping). He speculated that changes in irrigation practices in the Murray River area by the mid-1980s (specifically laser grading to minimise water pooling, and flood irrigation reduction) resulted in *Cx australicus* replacing *Ae. vigilax* as the primary initiating RRv vector in the area (although he also queried whether the lower than average spring and summer rainfall during the trapping study period may have abnormally reduced the abundance of *Aedes* species). *Cx australicus* dominated among the catches in Novembers and Decembers. Populations were observed to be plentiful at cool temperatures: substantial numbers were recorded at only 6°C, were most abundant between 17°C and 25°C, and declined rapidly from 32°C.

Russell (1993) has noted that *Cx australicus* does not normally bite humans, and feeds predominantly on rabbits and birds. If this species is capable of transmitting RRv, it is likely that it would not be involved in transmitting virus to humans (Lindsay *et al.* 1993b), but would be involved in the amplification of virus in the natural cycle.

The principal RRv transmission vector in the Murray River area is considered to be *Cx annulirostris*, a freshwater-breeding mosquito. Populations of female adult *Cx annulirostris* have been observed to overwinter in parts of the Murray Valley, and this is presumed to be the mechanism for their survival in the area (Russell 1986a). Dobrotworsky (1965) noted that *Cx annulirostris* generally breed from mid-spring to late-autumn in south-eastern Australia, with peak numbers appearing during late summer and in the autumn. Russell (1986b) conducted mosquito trapping around Echuca, Victoria, from 1979 to 1985. *Cx annulirostris* was recorded only between October and May (mid-spring to late-autumn). Numbers increased steadily to a peak in January or February, and then rapidly declined until April or May. *Cx annulirostris* was the dominant species in the area in the summer periods, comprising up to 92.4% of one year's sample, and on average more than 73% of samples from all years (Russell 1986b). McDonald (1980) reported a similar seasonal pattern around Mildura (located in Region 1) between 1975 and 1979, and *Cx annulirostris* comprised 16.7% to 90.3% of the summer mosquito populations. Numbers were observed to increase from late October and disappear by April or May.
Table 6.2  Breakdown of mosquito type from trapping conducted along the Murray River in Victoria, 1991/95. Table shows most abundant and consistently trapped species at ten sites over the four year period. Total mosquitoes trapped = 476 682.

<table>
<thead>
<tr>
<th>Species</th>
<th>% of total mosquitoes trapped</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cx annulirostris</td>
<td>60.6</td>
</tr>
<tr>
<td>Cx australicus</td>
<td>15.3</td>
</tr>
<tr>
<td>Ae. bancroftianus</td>
<td>5.0</td>
</tr>
<tr>
<td>Ae. sagax</td>
<td>3.3</td>
</tr>
<tr>
<td>An. annulipes</td>
<td>2.8</td>
</tr>
<tr>
<td>Cx pipiens</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Source: Dhileepan, 1996.

In the 1991-1995 study, Dhileepan (1996) also observed that Cx annulirostris was the most abundant species in the Murray Valley overall, followed by Cx australicus (see Table 6.2). Cx annulirostris was the dominant species from January onwards, increasing to a peak during February or March, with no activity recorded at all until November. This observation, and those of Russell (1986b) and McDonald (1980), supports the hypothesis that Cx annulirostris is not involved in the initiation and amplification of RRv in spring, but is responsible for transmission of the virus over summer.

Populations of Cx annulirostris have been observed to expand when maximum and minimum daily temperatures average 17.5°C, and to decline rapidly below this (McDonald 1980). Dhileepan (1996), however, observed that Cx annulirostris adults can occur at temperatures as low as 12°C, and Rae (1990) has reported 13.7°C to be the lower temperature threshold for adults. Populations have been found to peak when average temperatures reach 25°C (McDonald 1980, Mottram et al. 1986, Dhileepan 1996). Table 6.3 summarises the influence of temperature on Cx annulirostris larval development, population growth, and mosquito survival.

Rainfall in the Murray Valley has been positively related to Cx annulirostris population numbers, depending on the time of the year that it falls. Russell (1986b) trapped mosquitoes from 1979 to 1985, and observed that in drought years adult populations were very low (Russell 1986b). The highest mosquito numbers recorded followed regular substantial rainfall in spring and excessive rainfall in summer. Heavy rainfall at the end of a dry summer (in March and April), however, had almost no effect on low population numbers. Summer rainfall that produced flooding did not occur during the study, so the effect of extensive surface water on mosquito numbers was not recorded (Russell 1986b).
Table 6.3 Temperature influence on the larval development, population growth, and survival of *Culex annulirostris*.

<table>
<thead>
<tr>
<th>Life stage</th>
<th>Climate</th>
<th>Influence</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>larvae</td>
<td>mean daily temp = 25°C</td>
<td>optimal development</td>
<td>Lee <em>et al.</em>, 1989</td>
</tr>
<tr>
<td></td>
<td>mean daily temp = 30°C</td>
<td>optimal development</td>
<td>McDonald, 1980</td>
</tr>
<tr>
<td></td>
<td>mean daily temp &gt; 40°C</td>
<td>larval death</td>
<td>Lee <em>et al.</em>, 1989</td>
</tr>
<tr>
<td></td>
<td>mean daily temp &lt; 10°C</td>
<td>larval death</td>
<td>Lee <em>et al.</em>, 1989</td>
</tr>
<tr>
<td>adult</td>
<td>mean daily temp = 12°C</td>
<td>lower temperature threshold</td>
<td>Dhillapan, 1996</td>
</tr>
<tr>
<td></td>
<td>mean daily temp = 13.7°C</td>
<td>lower temperature threshold</td>
<td>Rae, 1990</td>
</tr>
<tr>
<td></td>
<td>mean daily temp &lt; 17.5°C</td>
<td>detrimental to population growth</td>
<td>McDonald, 1980</td>
</tr>
<tr>
<td></td>
<td>mean daily temp &gt; 35°C</td>
<td>detrimental to population growth</td>
<td>McDonald, 1980</td>
</tr>
<tr>
<td></td>
<td>mean daily temp 25-27°C</td>
<td>longest life span</td>
<td>Mottram, 1986</td>
</tr>
</tbody>
</table>

Substantial populations of *Culex* in spring (September to November) could increase virus activity, and have been suggested as a crucial factor in the generation of a summer epidemic (Russell 1986b).

Local observations in New South Wales from Mildura in the west to Albury in the east, and north to Griffith of weather patterns and RRv disease outbreaks (Terence Carvan pers. comm. 2000) have been that:

- High numbers of mosquitoes have been recorded when regular and substantial rain fell during the spring months (of sufficient quantity to lie in sheets on the ground, fill roadside channels, etc.).
- Excessive numbers of mosquitoes have been recorded when substantial rain also fell during the early summer months (i.e., December and January) when the average ambient temperatures were warmer.

Both these observations support the published literature of the influence of weather on the mosquito breeding patterns in the area.

### 6.3.2 Vertebrate hosts

The Western Grey kangaroo is common across the lower half of South Australia, and throughout western Victoria and New South Wales. The Red kangaroo is prevalent in the northern and western sections of the Murray Valley study area. The Eastern Grey kangaroo has a wide and almost continuous distribution between the inland plains and the eastern coast where the rainfall is more than 250 mm per annum (Strahan 1991), and
can be expected to be prevalent in the Murray River study areas. Russell (1995) has suggested that the Eastern Grey kangaroo is the dominant reservoir host in this area.

6.4 **Epidemiology of RRv Disease**

6.4.1 **Historical report of outbreaks**

Reports of the symptoms that came to be recognised as RRv disease in the Murray Valley area date from the period of European settlement in the late nineteenth century. The first documented outbreak of "epidemic polyarthritis" has generally been credited to Nimmo (1928). He observed "a number of cases of disease which [he could] not nail down" between March and April 1928 near Narrandera, New South Wales. That summer had been exceptionally wet, with considerable flooding (Nimmo 1946). Wolstenholme (1993), speculated that a report of cases of fever, rash, and severe aches and pains at Natimuk, western Victoria in 1885/86 (described by Weber 1886), initially diagnosed as dengue fever, may in fact have been an outbreak of epidemic polyarthritis. An unseasonal six inches of rain fell over a ten day period from late December to mid January, and cases were reported from late January.

Since then, several outbreaks have been reported in the Murray Valley. Widespread outbreaks, and some 200 cases, were reported in 1955/56 (Anderson and French 1957, Fuller and Warner 1957, Wilson 1957). Cases commenced in March, peaked in April, and waned in May-June (Fuller and Warner 1957). The summer of 1956 was a "particularly rainy season" (Fuller and Warner 1957), and April was a "month of maximum rainfall in an exceptionally wet season" (Wilson 1957). Excessive flooding, and the presence of large numbers of mosquitoes, was also observed (Anderson and French 1957, Wilson 1957).

Seglenieks and Moore (1974) reported an outbreak in 1971 that followed heavy rainfall and accompanied warm temperatures. Cases began in February, peaked in March, and dropped off in mid April - "paralleling the onset of wintry conditions". Cases were reported all along the Murray River in New South Wales, Victoria and South Australia. Rainfall for the area was generally above average from October 1970 through to June 1971; the Murray River had a high flow, and constant shallow flooding and mosquitoes were recorded.
In 1979/80, abnormally high late winter/early spring rainfall was recorded (November rainfall was 94% above average) (Mudge and Aaskov 1983). Large areas of water were observed to be lying on the ground (Mudge et al. 1980), and large populations of mosquitoes were reported. An outbreak of RRv disease commenced in late November, peaked between December and mid January, and waned in February.

In 1984, a very large outbreak of RRv disease occurred in summer throughout the whole of New South Wales (Hawkes et al. 1985). Cases in the Murray Valley commenced in approximately November 1983, and peaked in January. Abnormally heavy rainfall fell immediately before and during the outbreak, and a prolonged drought preceded it.

<table>
<thead>
<tr>
<th>Year</th>
<th>Winter</th>
<th>Spring</th>
<th>Summer</th>
<th>Autumn</th>
</tr>
</thead>
<tbody>
<tr>
<td>1886</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1928</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1971</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.4 The seasonal pattern of historic RRv disease epidemics and rainfall in the Murray Valley area, summarised from the literature.

This history is summarised graphically in Table 6.4. All documented large outbreaks in the Murray Valley have been preceded by or have occurred during above average rainfall (with widespread flooding noted in many accounts). Outbreaks have commenced either in late spring/early summer, or in late summer/autumn (the warmer months of the year). Two epidemic patterns are suggested from this history:

1. High rainfalls in late winter or spring, followed by cases appearing three to four months later (1970/71, 1979/80). Additional heavy rainfall throughout summer may or may not have occurred.

2. Heavy rainfalls in early summer, followed by cases from two weeks to less than four weeks afterwards (1885/86, 1927/28, 1955/56, 1983/84).

6.4.2 Study period
For Region 1, 1289 notifications of RRv disease were recorded from July 1991 to June 1999 (Figure 6.5 is a graph of notifications for each region). There were less male notifications than female ones (ratio 1:1.22), whereas the national sex ratio was almost equal (1:1.03). The majority of notifications were in the 25-59 year age group (72%), with only 4% of notifications recorded in children aged less than 15 years (essentially equivalent to the national population). The eight-year average annual incidence for the 14 Region 1 SLAs was 142/100 000 (range 75-1126/100 000), almost six times higher than the national average rate for the period (25/100 000).

For Region 2, 809 notifications were recorded for the period (Figure 6.5). There were more male notifications than female ones (ratio 1:0.83). The majority (78%) of notifications were in the 25-59 year age group, and 3% of notifications were of children aged less than 15 years. The eight-year average annual incidence rate for the 24 Region 2 SLAs was very much lower than for Region 1: 13/100 000 (range 4-131/100 000), and lower than the national average (25/100 000).

Table 6.5 gives the number of epidemics in each study region per year. In any given year, a maximum of 14 epidemics or non-epidemics could have occurred in Region 1 (i.e., one per SLA). In Region 2 a maximum of 24 epidemic or non-epidemics could have occurred.

<table>
<thead>
<tr>
<th>Year</th>
<th>Region 1</th>
<th>Region 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>92/93</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>93/94</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>94/95</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>95/96</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>96/97</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>97/98</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>98/99</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>86</td>
</tr>
</tbody>
</table>

* The definition of an epidemic used for the Murray study area was: the mean plus one standard deviation of all the cases recorded during the study period (Chapter 5).
† Number of epidemics|non-epidemics in a region during the period.
Figure 6.5  RRv disease notifications for each Murray study region, by year (July 1991 to June 1999).
Figure 6.6  RRv disease notifications for Region 1 and Region 2 in the Murray, by month of onset (July 1991-June 1999).

Murray Region 1

Murray Region 2
Widespread epidemics occurred in 1992/93 (Region 1: 12/14 SLAs; Region 2: 24/24 SLAs) and in 1996/97 (Region 1: 11/14 SLAs; Region 2: 19/24 SLAs). No epidemics were recorded in the years 1993/94 to 1995/96 inclusive. Few epidemics were recorded in SLAs in the other years.

The pattern of monthly disease notifications was very similar for both regions. Figure 6.6 shows the average number of notifications for both regions by the estimated month of RRv disease onset. The majority of notifications were recorded from January to April in both regions, with the peak percentage in February (Region 1, 31%; and Region 2, 30%). For the period overall, 54% of notifications had been recorded by the end of February in Region 1, and 58% in Region 2. The pattern of cases peaking in February was evident for all epidemic years, except in Region 2 in 1992/93. In that year epidemics occurred in all SLAs, and the number of cases peaked in March.

In 1992/93 in Region 1, the first cases were notified in September (in 1 of 12 SLAs), in November (in 1 of 12 SLAs), in January (in 10 of 12 SLAs). In 1996/97, the first cases were notified in November (in 1 of 11 SLAs), December (in 6 of 11 SLAs), in January (in 4 of 11 SLAs). In 1992/93 in Region 2, the first cases were notified in November (in 1 of 24 SLAs), in December (in 1 of 24 SLAs), in January (in 10 of 24 SLAs), in February (in 7 of 24 SLAs), in March (in 5 of 24 SLAs). In 1996/97, the first cases were notified in December (in 3 of 19 SLAs), in January (in 9 of 19 SLAs), in February (in 4 of 19 SLAs), and in March (in 3 of 19 SLAs).

6.5 Hypotheses

The main thesis question that guided this study was “Can epidemics of RRv disease be predicted with sufficient accuracy, timing and reliability to be of use to public health authorities for early warning and response?”

Several specific hypotheses were developed to guide the development of modelling in the Murray study area. Based on the material presented in the previous sections, the following conclusions were drawn:

- First, a positive relationship has been noted between high rainfall in the spring to summer (i.e., warm) months of the preceding year and high mosquito numbers. High
mosquitoes numbers in these months would presumably promote circulation of virus between the vector and host populations. This in turn would lead to a reduction in the total number of susceptible (i.e., non-immune) vertebrate hosts. Conditions for human epidemics require high mosquito numbers, plus high levels of viral activity (i.e., low levels of immunity in the vertebrate host population). The primary hosts (kangaroos) take more than one full year to complete their breeding cycle, and thus fewer susceptible hosts may be expected for the year following heavy spring or summer rainfalls.

- Second, *Cx australicus*, and or *Ae. vigilax*, appear to be responsible for initiating and amplifying the virus in the vertebrate host population from at least October and possibly in earlier months. If *Cx australicus* was the dominant species, less extreme temperatures in late spring appear most conducive to breeding (between 6 and 25°C, with the optimum at 20°C). Its ability to breed at lower temperatures means that late winter to late spring rainfall may be an important factor in sustaining early population growth.

- Third, the majority (but not all) of past recorded epidemics were preceded by or accompanied above average summer rainfall.

- Fourth, the extrinsic incubation period for mosquitoes decreases substantially as the weather warms (although extreme temperatures can lead to mosquito death). High summer temperatures may be most suitable for an extended survival period for *Cx annulirostris*, thus increasing the risk of RRv infection of human populations.

Drawing on these points, several hypotheses were examined:

1. That rainfall in the late spring to summer months of the lagged year (i.e., the year preceding the epidemic year) is negatively associated with epidemics.
2. That rainfall in late winter to early spring is positively associated with epidemics.
3. That surrogate indicators of rainfall in the same months – the SOI and SST – are positively and negatively associated with epidemics respectively.
4. That late spring temperature is negatively associated with epidemics.
5. That summer rainfall is positively associated with epidemics.
6. That summer temperature is positively associated with epidemics.
7. That extreme summer temperature is negatively associated with epidemics.
6.6 METHODS

The methods for the predictive modelling are covered in Chapter 5. Briefly, single variable regression analysis was conducted to identify candidate variables. With the candidate variables for months 1-5, multivariable logistic regression was used to develop an early warning model for each region. A late warning prediction model was also developed for each region, using variables for months 1-8. Best-fitting models were chosen according to the established criteria. The models were cross-validated within the data set, and the ability of the models to predict epidemics in individual years and for all years was calculated. In addition, data from the two regions were pooled, and a "combined model" was built for the early warning and late warning periods, to test the usefulness of the RRv bioclimatic regions as a method for defining the boundaries of analysis.

Note that a "year" in the context of this thesis extends from the month of July to the month of June. Thus, July=month 1, August=month 2, etc. Also note that the seasons reported are for the Southern Hemisphere: winter (June to August); autumn (September to November); summer (December to February); spring (March to May).

6.7 RESULTS

This section describes the results of the logistic regression predictive modelling for both regions.

6.7.1 Single variable analysis

Single variable logistic regressions were used to examine the relationship between the climatic ("explanatory") variables and the occurrence of an epidemic (the "outcome variable"). As discussed in Chapter 5, lags were tested for the rainfall variables (total rainfall, number of rain days, SOI, and SST). Only significant results are reported.
6.7.1.1 Effect of rainfall

Figures 6.7 and 6.8 show the spatially averaged rainfall patterns in the epidemic years for both regions for the period. As expected from the long-term average record, greater quantities fell in almost every month in Region 2 (temperate) than in Region 1 (semi-arid). In the 1992/93 epidemic year, total rainfall throughout the late winter and early spring months (July to September) was 47% higher than the long-term average across Region 1, and 37% higher in Region 2. Summer rainfall in the same year (December to February) was much higher than the long-term average (+134% in Region 1, and +107% in Region 2). In the 1996/97 epidemic year, late winter and early spring rainfall was also higher than the average in both regions (+35% in Region 1, and +37% in Region 2). However, summer rainfall was under average values (-36% in Region 1, and -8% in Region 2).

In both regions, the two precipitation variables (number of rain days and total rainfall) in the late winter and early spring months were strongly and positively associated with epidemics (Figures 6.9 and 6.10). In Region 1, rain days and total rainfall in August and September were most strongly associated. In Region 2, total rainfall in August and September, and rain days in September were the strongest predictors.

Derived variables

To test the hypothesis that rainfall in the preceding year was negatively associated with epidemics, I derived lags of the rainfall variables (total rainfall and rain days). One year lags for the spring and summer months were tested separately against the occurrence of an epidemic, and averages of months were also tested.

For both regions a statistically significant negative relationship was observed between epidemics and one year lagged rain days for the months of September through to December. The strongest relationship was an average of lagged rain days for October to December. The measures of association for Region 1 were: OR=0.59, CI=0.40-0.85; p=0.004. For Region 2 they were: OR=0.72, CI=0.61-0.85; p=0.000. The probability of an epidemic increased substantially when the number of rain days in the three month period was exceptionally low (less than 3-4).
Figure 6.7 1992/93 epidemic year: mean total rainfall for the two Murray study regions, by month.

Figure 6.8 1996/97 epidemic year: mean total rainfall for the two Murray study regions, by month.
**Figure 6.9** Odds ratios for rainfall regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at $P<0.05$.

**Figure 6.10** Odds ratios for the number of rain days regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at $P<0.05$. 
Figure 6.11 Odds ratios for the Southern Oscillation Index regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at $P<0.05$.

Region 1

Region 2

Figure 6.12 Odds ratios for Sea Surface Temperature regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at $P<0.05$. 
Figure 6.13  Odds ratios for average minimum temperature regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at P<0.05.

![Figure 6.13](image)

Figure 6.14  Odds ratios for absolute minimum temperature regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at P<0.05.

![Figure 6.14](image)
Figure 6.15  Odds ratios for average temperature regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at P<0.05.

Region 1

Region 2

Figure 6.16  Odds ratios for average maximum temperature regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at P<0.05.
Figure 6.17  Odds ratios for absolute maximum temperature regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at $P<0.05$.

Figure 6.18  Odds ratios for relative humidity at maximum temperature regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at $P<0.05$. 
**Figure 6.19** Odds ratios for relative humidity at minimum temperature regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at $P<0.05$.

**Figure 6.20** Odds ratios for vapour pressure regressed against epidemics of RRv disease in the Murray regions. Odds ratios on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at $P<0.05$. 
6.7.1.2  **Effect of the SOI and SST**

The SOI for the months of August and September was a significant positive predictor of epidemics in both regions (Figure 6.11). Region 1, measures of association for the August SOI (OR=1.15, CI=1.08-1.23, P<0.000) and September SOI (OR=1.12, CI=1.06-1.19, P<0.000). Region 2, measures of association for the August SOI (OR=1.13, CI=1.08-1.18, P<0.000) and September SOI (OR=1.11, CI=1.06-1.16, P<0.000).

SST for the months of September, October and November were significant negative predictors of epidemics for both regions (Figure 6.12). Region 1 SST September (OR=0.46, CI=0.24-0.91, P<0.025), SST October (OR=0.40, CI=0.24-0.69, P<0.001), and SST November (OR=0.40, CI=0.25-0.66, P<0.000). Region 2 SST September (OR=0.63, CI=0.40-0.97, P<0.037) and SST October (OR=0.45, CI=0.31-0.66, P<0.000), SST November (OR=0.44, CI=0.31-0.63, P<0.000).

6.7.1.3  **Effect of temperature**

Moderate temperatures in winter and spring were strongly related to the occurrence of epidemics. In late winter/spring (month 2=August to month 5=November), high minimum temperatures (Figures 6.13 and 6.14) and low average and maximum temperatures (Figures 6.15 and 6.17) were associated with the occurrence of epidemics. In Region 1 (semi-arid), 96% of epidemics occurred when November maximum temperatures were 26°C or less, compared to 50% of non-epidemics. In Region 2 (closer to the coast), 91% of epidemics occurred when November maximum temperatures were 22°C or less, compared to 31% of non-epidemics.

All February temperature variables (late summer) were positively associated with epidemics in both regions. Odds ratios for the temperature variables in February were in all cases higher in Region 1 (semi-arid) than in Region 2 (temperate). The average temperature range is more extreme in Region 1, and an increase in temperatures there would be likely to reduce the extrinsic incubation period (and increase the risk of transmission) more than in Region 2.
Chapter 6 - The Murray River study area

Derived variables

I tested whether there was a threshold effect for the relationship between minimum temperatures and epidemics in the late winter months, which might interrupt *Aedes* breeding and affect the early initiation of virus in the regions. Results were significant, but the width of the confidence intervals for the odds ratios suggested that the prediction was not reliable. I also tested whether high late spring temperatures had a threshold effect on the build-up of mosquito number. Again, this increased the odds ratio, but the wide confidence intervals indicated the time series was not long enough to reliably estimate the coefficient for this variable.

The interaction between temperature and relative humidity and evaporation was examined in an additive and multiplicative fashion to determine whether these factors would better predict epidemics than temperature on its own. No significant relationships were found. Variables were also entered as squared and cubed values, to test whether these could better approximate the non-linear relationship with mosquito breeding. No substantial improvement in prediction occurred.

6.7.1.4  Effect of humidity

There was a positive association between relative humidity (calculated at maximum temperature) and epidemics for most months except July and February in both regions. The strongest relationship in Region 1 was in August (OR=1.32, CI=1.15-1.52, p<0.00), and in September in Region 2 (OR=1.48, CI=1.30-1.69, p<0.00). See Figure 6.18 for a graph of the odds ratios.

The association between relative humidity calculated at minimum temperatures and epidemics was non-significant in most months (Figure 6.19). Positive associations in both regions were observed for October (Region 1: OR=1.20, CI=1.07-1.34, p<0.001; Region 2: OR=1.12, CI=1.04-1.21, p<0.002) and November (Region 1: OR=1.11, CI=1.04-1.19, p<0.002; Region 2: OR=1.05, CI=1.00-1.02, p<0.048).

Vapour pressure was positively associated with epidemics during all summer months in both regions (Figure 6.20).
6.7.2 Early warning models (months 1-5)

Multivariable logistic regression was used to examine the relationship between climate and epidemics. The aim was to see whether a model could be developed for each region that was (i) accurate, (ii) reliable, and (iii) that could provide enough lead time for public health response to an epidemic. In recognition of the third aim, the early warning models comprised explanatory variables only for the first five months of the year (July=1 through to November=5). The method for building the models is described fully in Chapter 5.

6.7.2.1 Region 1 early warning model

Epidemics in this region were best predicted by (i) the number of rain days in the spring months, lagged by one year, (ii) early spring rainfall in the current year, and (iii) late spring temperature variables in the current year. The logistic regression model with the best fit had the form:

\[
\log \left( \frac{P}{1-P} \right) = a + \beta_1 \times (\text{lagRainDays}_{456}) + \beta_2 \times (\text{RainDays}_{1}) + \beta_3 \times (\text{Rain}_{23}) + \beta_4 \times (\text{Humid}_{5}) + \beta_5 \times (\text{MaxTmp}_{5})
\]

Table 6.6 shows the predictor variables, their coefficient estimates, standard errors, odds ratios, and 95% confidence intervals. All variables were significant at the P<0.05 level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>P value</th>
<th>Coeffs</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>One year lag rain days (Oct-Dec)</td>
<td>0.25</td>
<td>0.07 - 0.91</td>
<td>0.035</td>
<td>-1.37</td>
<td>0.65</td>
</tr>
<tr>
<td>Rain days (July)</td>
<td>2.92</td>
<td>1.21 - 7.03</td>
<td>0.017</td>
<td>1.07</td>
<td>0.45</td>
</tr>
<tr>
<td>Rainfall (Aug-Sep)</td>
<td>1.87</td>
<td>1.02 - 3.42</td>
<td>0.043</td>
<td>0.63</td>
<td>0.31</td>
</tr>
<tr>
<td>Relative humidity at minimum temperature (Nov)</td>
<td>0.27</td>
<td>0.09 - 0.81</td>
<td>0.020</td>
<td>-1.30</td>
<td>0.56</td>
</tr>
<tr>
<td>Average maximum temperature (Nov)</td>
<td>0.01</td>
<td>0.00 - 0.37</td>
<td>0.012</td>
<td>-4.55</td>
<td>1.81</td>
</tr>
</tbody>
</table>

* Model R² (percentage of variance explained by model) = 84%

Rain days for October to December of the year preceding the epidemic year were negatively associated with epidemics. The number of rain days, rather than total rainfall for this period, was a better predictor. In the year prior to the epidemics of 1992/93, the average number of rain days from October to December was extremely low (2.0 days recorded for the whole three months). In the year before the 1996/97 epidemics, the number of rain days was 4.2 (range of values for the period was 1-14).
Chapter 6 - The Murray River study area

The number of rain days in July and average rainfall for August and September were significant positive predictors of epidemics. The SST and SOI variables for the late winter to spring months, although significant in the single variable analysis, were highly correlated with rainfall and were dropped from the model.

All average and maximum November temperature variables were negatively associated with epidemics in Region 1. Mean maximum temperature was a better predictor than any of the others. Only 4% of epidemics occurred in years when November mean maximum temperatures exceeded 26°C, compared to 50% of non-epidemics.

Relative humidity at minimum temperatures in late spring was a negative predictor of epidemics. Other temperature-related variables that were significant in single variable analysis (such as October vapour pressure) were correlated with November relative humidity and maximum temperature, and were dropped from the model.

Model fit

The R² value (84%) for the model indicates that a large proportion of the variation in the outcome was explained. The Hosmer and Lemeshow goodness-of-fit test was 0.5 (prob > chi² = 0.999), which suggests that the model fit the data very well. The Stata diagnostics, however, indicated that some 5% of observations were not well explained. For the unvalidated model, 91% of observations were correctly fitted by the model (90% was the specified criterion, outlined in Chapter 5).

Cross-validation

Table 6.7 shows the results of the validation test for individual years.³ The model was better at correctly identifying non-epidemic years than epidemic ones (overall specificity of 95%, with most years at 100%). For the whole period, 62% of epidemics were correctly identified. In the years when the majority of epidemics occurred (92/93 and 96/97) 55-58% were identified. Related to this, the model had very low negative predictive ability in

---

³ Validation Test. Sensitivity: the proportion of epidemics that the model correctly identified as epidemics. Specificity: the proportion of non-epidemics that the model correctly identified as non-epidemics. Positive predictive value: the likelihood that an epidemic predicted by the model really is an epidemic. Negative predictive value: the likelihood that a non-epidemic predicted really is a non-epidemic. Accuracy: the percentage of epidemics and non-epidemics correctly predicted. Except where otherwise noted, a cut-off probability of 0.5 was used to define an epidemic.
the epidemic years (i.e., there was low confidence that when no epidemic was predicted it really did not occur).

Table 6.7  Sensitivity, specificity, positive predictive values, and negative predictive values for the early warning model in Murray Region 1, by year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>100%</td>
<td>69%</td>
<td>20%</td>
<td>100%</td>
</tr>
<tr>
<td>92/93</td>
<td>58%</td>
<td>100%</td>
<td>100%</td>
<td>29%</td>
</tr>
<tr>
<td>93/94</td>
<td>–</td>
<td>100%</td>
<td>–</td>
<td>100%</td>
</tr>
<tr>
<td>94/95</td>
<td>–</td>
<td>100%</td>
<td>–</td>
<td>100%</td>
</tr>
<tr>
<td>95/96</td>
<td>–</td>
<td>100%</td>
<td>–</td>
<td>100%</td>
</tr>
<tr>
<td>96/97</td>
<td>55%</td>
<td>100%</td>
<td>100%</td>
<td>38%</td>
</tr>
<tr>
<td>97/98</td>
<td>–</td>
<td>100%</td>
<td>–</td>
<td>100%</td>
</tr>
<tr>
<td>98/99</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>All years</td>
<td>62%</td>
<td>95%</td>
<td>80%</td>
<td>89%</td>
</tr>
</tbody>
</table>

The percentage of epidemics and non-epidemics that were accurately predicted (Table 6.8) did not vary significantly as the threshold level for determining the probability of an epidemic was raised. At the 0.50, 0.70, and 0.90 cut-offs it was around 90% for all years, which was surprisingly high.

Table 6.8  Accuracy of the early warning model for Murray Region 1, using varying cut-off values to differentiate the probability of an epidemic, by year.

<table>
<thead>
<tr>
<th>Year</th>
<th>0.5 cut-off</th>
<th>0.7 cut-off</th>
<th>0.9 cut-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>71%</td>
<td>93%</td>
<td>100%</td>
</tr>
<tr>
<td>92/93</td>
<td>64%</td>
<td>64%</td>
<td>50%</td>
</tr>
<tr>
<td>93/94</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>94/95</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>95/96</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>96/97</td>
<td>64%</td>
<td>64%</td>
<td>64%</td>
</tr>
<tr>
<td>97/98</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>98/99</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>All years</td>
<td>88%</td>
<td>90%</td>
<td>89%</td>
</tr>
</tbody>
</table>
6.7.2.2 Region 2 early warning model

The model of best fit for Region 2 was very similar in generalities to Region 1. Epidemics were also best predicted by (i) the number of rain days in spring lagged by one year, (ii) early spring rainfall, and (iii) late spring temperature variables. The logistic regression model with the best fit had the form:

\[
\log\left(\frac{P}{1-P}\right) = \alpha + \beta_1(\text{lagRainDays}_{456}) + \beta_2(\text{RainDays}_1) + \beta_3(\text{SST}_2) + \beta_4(\text{RainDays}_3) + \beta_5(\text{MaxTmp}_5)
\]

Table 6.9 shows the predictor variables, their coefficient estimates, standard errors, odds ratios, and 95% confidence intervals. All variables were significant at the P<0.05 level.

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>P value</th>
<th>Coeffs</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>One year lag rain days (Oct-Dec)</td>
<td>0.36</td>
<td>0.22 - 0.59</td>
<td>0.000</td>
<td>-1.03</td>
<td>0.09</td>
</tr>
<tr>
<td>Rain days (July)</td>
<td>1.4</td>
<td>1.10 - 1.78</td>
<td>0.006</td>
<td>0.34</td>
<td>0.17</td>
</tr>
<tr>
<td>Sea surface temperature (Aug)</td>
<td>27.9</td>
<td>5.05 - 154</td>
<td>0.024</td>
<td>3.33</td>
<td>24.38</td>
</tr>
<tr>
<td>Rain days (Sept)</td>
<td>1.49</td>
<td>1.05 - 2.12</td>
<td>0.000</td>
<td>0.40</td>
<td>0.27</td>
</tr>
<tr>
<td>Absolute maximum temperature (Nov)</td>
<td>0.37</td>
<td>0.24 - 0.57</td>
<td>0.000</td>
<td>-0.99</td>
<td>0.08</td>
</tr>
</tbody>
</table>

* Model R² (percentage of variance explained by model) = 77%

As in Region 1, the number of rain days in October to December of the previous year was negatively associated with epidemics (and was a better predictor than the total rainfall variables for the same period). In the year before the epidemics of 1992/93, the number of rain days for October to December was low (3.4 days recorded for the three months). In the year before the 1996/97 epidemics, the number of rain days for those months was 5.4 (range of values for the period 1-14).

Rainfall well above average was sustained from June to September 1992/93. The precipitation variables that best predicted epidemics were the number of rain days in July and September, and August sea surface temperature (which was a better predictor than the SOI). The confidence intervals for the odds ratios for August SST were wide, reflecting the narrow range of values available for estimation.

November temperatures were also strongly (and negatively) associated with epidemics in Region 2, and absolute maximum temperature was the best predictor. Only 6% of
epidemics occurred in years when absolute maximum temperature exceeded 32°C, compared to 80% of non-epidemics.

Model fit

The $R^2$ value (78%) for the model indicates that a high proportion of the variation in the outcome was explained. The Hosmer and Lemeshow goodness-of-fit test was 2.05 (prob $>\text{chi}^2 = 0.979$), which suggests that the model fitted the data very well. For the unvalidated model, 92% of observed values were correctly fitted.

Cross-validation

The results of the validation test for individual years are provided in Table 6.10. The sensitivity for all years (73%) was higher than in Region 1, and the specificity was lower (81%). The model sensitivity for the two main epidemic years was between 63% and 79%. The specificity was very poor in 1997/98: the model incorrectly predicted that epidemics would occur in all SLAs (although only one did). There was a correspondingly low positive predictive value for that year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>92/93</td>
<td>79%</td>
<td>-</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>93/94</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>94/95</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>95/96</td>
<td>-</td>
<td>88%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>96/97</td>
<td>63%</td>
<td>60%</td>
<td>86%</td>
<td>30%</td>
</tr>
<tr>
<td>97/98</td>
<td>100%</td>
<td>0%</td>
<td>4%</td>
<td>-</td>
</tr>
<tr>
<td>98/99</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>All years</td>
<td>73%</td>
<td>81%</td>
<td>53%</td>
<td>91%</td>
</tr>
</tbody>
</table>

Table 6.10 Sensitivity, specificity, positive predictive values, and negative predictive values for the early warning model in Murray Region 2, by year.

---

6 Validation Test. Sensitivity: the proportion of epidemics that the model correctly identified as epidemics. Specificity: the proportion of non-epidemics that the model correctly identified as non-epidemics. Positive predictive value: the likelihood that an epidemic predicted by the model really is an epidemic. Negative predictive value: the likelihood that a non-epidemic predicted really is a non-epidemic. Accuracy: the percentage of epidemics and non-epidemics correctly predicted. Except where otherwise noted, a cut-off probability of 0.5 was used to define an epidemic.
Chapter 6 - The Murray River study area

The percentage of epidemics and non-epidemics that were accurately predicted decreased slightly as the threshold level for determining the probability of an epidemic was raised (Table 6.11). At a 0.50 cut-off, the accuracy of the model was 79%, at a 0.70 cut-off it was 77%, and at a 0.90 cut-off it was 71%.

<table>
<thead>
<tr>
<th>Year</th>
<th>0.5 cut-off</th>
<th>0.7 cut-off</th>
<th>0.9 cut-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>92/93</td>
<td>79</td>
<td>67</td>
<td>38</td>
</tr>
<tr>
<td>93/94</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>94/95</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>95/96</td>
<td>88</td>
<td>92</td>
<td>96</td>
</tr>
<tr>
<td>96/97</td>
<td>63</td>
<td>50</td>
<td>33</td>
</tr>
<tr>
<td>97/98</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>98/99</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>All years</td>
<td>79</td>
<td>77</td>
<td>71</td>
</tr>
</tbody>
</table>

6.7.3 Late warning models (months 1-8)

The late warning models included explanatory variables for the first eight months of the year (July=1 to February=8). February is, technically, the last month of the Austral summer.

6.7.3.1 Region 1 late warning model

In addition to:
- lagged spring rain days,
- early spring rainfall, and
- late spring temperatures,
late summer temperatures were significant predictors in the Region 1 model, and made a striking improvement to the predictive accuracy of the Region 1 model. The logistic regression model of best fit had the form:

$$\log\left[\frac{P}{1-P}\right] = a + \beta_1(\text{lagRainDays}_{456}) + \beta_2(\text{RainDays}_{1}) + \beta_3(\text{Rain}_{23}) + \beta_4(\text{MaxTmp}_{5}) + \beta_5(\text{Humid}_{8}) + \beta_6(\text{MinTmp}_{8})$$
Table 6.12 shows the predictor variables, their coefficient estimates, standard errors, odds ratios, and 95% confidence intervals. All variables were significant at the $P<0.05$ level.

**Table 6.12** Odds ratios (OR), 95% confidence intervals (CI), coefficients, and standard errors (SE) for the late warning logistic regression model in Region 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>P value</th>
<th>Coeffs</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>One year lag rain days (Oct-Dec)</td>
<td>0.15</td>
<td>0.03 - 0.78</td>
<td>0.024</td>
<td>-1.89</td>
<td>0.84</td>
</tr>
<tr>
<td>Rain days (July)</td>
<td>1.88</td>
<td>1.1 - 3.3</td>
<td>0.029</td>
<td>0.63</td>
<td>0.29</td>
</tr>
<tr>
<td>Total rainfall (Aug-Sept)</td>
<td>1.34</td>
<td>1.1 - 1.6</td>
<td>0.004</td>
<td>0.29</td>
<td>0.10</td>
</tr>
<tr>
<td>Absolute maximum temperature (Nov)</td>
<td>0.37</td>
<td>0.2 - 0.8</td>
<td>0.007</td>
<td>-0.99</td>
<td>0.37</td>
</tr>
<tr>
<td>Relative humidity at maximum temperature (Feb)</td>
<td>0.25</td>
<td>0.1 - 0.8</td>
<td>0.016</td>
<td>-1.40</td>
<td>0.58</td>
</tr>
<tr>
<td>Absolute minimum temperature (Feb)</td>
<td>4.19</td>
<td>1.2 - 14.0</td>
<td>0.021</td>
<td>1.43</td>
<td>0.62</td>
</tr>
</tbody>
</table>

* Model $R^2$ (percentage of variance explained by model) = 79%

One year lagged spring rain days in July, and rainfall for August to September, were predictor variables in the late warning model, as they were in the early warning model. However, instead of mean maximum temperatures for November (one of the early warning model variables for this region), absolute maximum temperatures contributed more to the late warning model.

Relative humidity (calculated at maximum temperature) in February was negatively associated with epidemics. Absolute minimum temperatures in the relative humidity at maximum temperature were positively associated with epidemics, although the confidence interval was somewhat wide (the range of values available to estimate the coefficient of this variable was narrow: $6^\circ\text{C}$ to $13^\circ\text{C}$).

A very different summer rainfall pattern between the two epidemic years prevented any of the rainfall variables for that season from contributing to the model (compare the rainfall values for the December to February months in Figures 6.7 and 6.8).

**Model fit**

The $R^2$ value for the model (79%) indicated that the model explained a reasonable proportion of the variation in the outcome. The Hosmer and Lemeshow goodness-of-fit test was 26.36 (prob $>\chi^2$ = 0.09), which suggests that the model had problems fitting the data. For the unvalidated model, 91% of observed values were correctly fitted.
Cross-validation

The results of the validation test are provided in Table 6.13. The sensitivity for all years (96%) was very high, as was the specificity (93%). The model had a tendency to misclassify non-epidemics as epidemics (positive predictive value = 81%).

<table>
<thead>
<tr>
<th>Year</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>0%</td>
<td>100%</td>
<td>-</td>
<td>93%</td>
</tr>
<tr>
<td>92/93</td>
<td>100%</td>
<td>50%</td>
<td>92%</td>
<td>100%</td>
</tr>
<tr>
<td>93/94</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>94/95</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>95/96</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>96/97</td>
<td>100%</td>
<td>33%</td>
<td>85%</td>
<td>100%</td>
</tr>
<tr>
<td>97/98</td>
<td>-</td>
<td>86%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>98/99</td>
<td>100%</td>
<td>92%</td>
<td>67%</td>
<td>100%</td>
</tr>
<tr>
<td>All years</td>
<td>96%</td>
<td>93%</td>
<td>81%</td>
<td>99%</td>
</tr>
</tbody>
</table>

The percentage of epidemics and non-epidemics that were accurately predicted by the model decreased only little as the threshold level for determining the probability of an epidemic was raised (Table 6.14). At a 0.5 cut-off the accuracy was 94%, at a 0.7 cut-off it was 89%, and at a 0.9 cut-off it was 88%. However, the accuracy in the 1992/93 epidemic year reduced more substantially as the cut-off was raised (from 93% to 36%). This indicates that the model produced more marginal probabilities for an epidemic in that year.

---

7 Validation Test. Sensitivity: the proportion of epidemics that the model correctly identified as epidemics. Specificity: the proportion of non-epidemics that the model correctly identified as non-epidemics. Positive predictive value: the likelihood that an epidemic predicted by the model really is an epidemic. Negative predictive value: the likelihood that a non-epidemic predicted really is a non-epidemic. Accuracy: the percentage of epidemics and non-epidemics correctly predicted. Except where otherwise noted, a cut-off probability of 0.5 was used to define an epidemic.
### Table 6.14
Accuracy of the late warning model for Murray Region 1, using varying cut-off values to differentiate the probability of an epidemic, by year.

<table>
<thead>
<tr>
<th>Year</th>
<th>0.5 cut-off</th>
<th>0.7 cut-off</th>
<th>0.9 cut-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>93</td>
<td>93</td>
<td>93</td>
</tr>
<tr>
<td>92/93</td>
<td>93</td>
<td>50</td>
<td>36</td>
</tr>
<tr>
<td>93/94</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>94/95</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>95/96</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>96/97</td>
<td>86</td>
<td>86</td>
<td>86</td>
</tr>
<tr>
<td>97/98</td>
<td>86</td>
<td>86</td>
<td>100</td>
</tr>
<tr>
<td>98/99</td>
<td>93</td>
<td>100</td>
<td>86</td>
</tr>
<tr>
<td>All years</td>
<td>94</td>
<td>89</td>
<td>88</td>
</tr>
</tbody>
</table>

### 6.7.3.2 Region 2 late warning model

In addition to:
- lagged spring rain days,
- early spring rainfall, and
- late spring temperatures,
late summer temperatures were also significant predictors in Region 2. The logistic regression model that provided the best fit had the form:

\[
\log\left[\frac{P}{1-P}\right] = \alpha + \beta_1 \cdot (\text{lagRainDays}_{456}) + \beta_2 \cdot (\text{RainDays}_1) + \beta_3 \cdot (\text{MaxTmp}_5) + \beta_4 \cdot (\text{MaxTmp}_6) + \beta_5 \cdot (\text{VapourPress}_8) + \beta_6 \cdot (\text{MaxTmp}_8)
\]

Table 6.15 gives the predictor variables, their coefficient estimates, standard errors, odds ratios, and 95% confidence intervals. All variables were significant at the P<0.05 level.

### Table 6.15
Odds ratios (OR), 95% confidence intervals (CI), coefficients, and standard errors (SE) for the late warning logistic regression model for Region 2.*

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>P value</th>
<th>Coeffs</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>One year lag rain days (Oct-Dec)</td>
<td>0.60</td>
<td>0.41 – 0.89</td>
<td>0.011</td>
<td>-0.51</td>
<td>0.20</td>
</tr>
<tr>
<td>Rain days (Sept)</td>
<td>1.52</td>
<td>1.02 – 2.27</td>
<td>0.041</td>
<td>0.42</td>
<td>0.21</td>
</tr>
<tr>
<td>Mean maximum temperature (Nov)</td>
<td>0.45</td>
<td>0.20 – 0.99</td>
<td>0.047</td>
<td>-0.80</td>
<td>0.40</td>
</tr>
<tr>
<td>Absolute maximum temperature (Dec)</td>
<td>0.38</td>
<td>0.19 – 0.77</td>
<td>0.007</td>
<td>-0.96</td>
<td>0.35</td>
</tr>
<tr>
<td>Vapour pressure (Jan-Feb)</td>
<td>5.73</td>
<td>1.46 – 22.4</td>
<td>0.001</td>
<td>1.74</td>
<td>0.70</td>
</tr>
<tr>
<td>Mean maximum temperature (Feb)</td>
<td>2.14</td>
<td>1.38 – 3.31</td>
<td>0.012</td>
<td>0.76</td>
<td>0.22</td>
</tr>
</tbody>
</table>

* Model $R^2$ (percentage of variance explained by model) = 80%
The odds ratio for rain days in September was almost unchanged in this model (compared to the early warning model). The two other rainfall variables for the late winter to early spring period from the early warning model were dropped from this late warning model.

As in Region 1, the summer rainfall variables were not predictors in the late warning model.

Low temperatures in the early summer were predictive of epidemics, continuing the pattern from the early warning model. Average February temperature was positively associated with epidemics (and was a better predictor in this region than minimum temperatures in the same month). Vapour pressure in mid to late summer was positively associated with epidemics, although the confidence intervals for this variable were wide (1.46-22.4).

**Model fit**

The $R^2$ value (80%) suggests that a reasonable proportion of the variation in the outcome was explained by the model. The Hosmer and Lemeshow goodness-of-fit test was 0.98 (prob $>\chi^2 = 0.998$), which suggests that the model fitted the data well. For the unvalidated model, 92% of observed values were correctly fitted.

**Cross-validation**

The results of the validation test are provided in Table 6.16. The overall sensitivity (66%) was lower than for the early warning model (which was 73%). All epidemics that occurred in 1992/93 were correctly predicted. However, only 26% of epidemics in 1996/97 were predicted. 1997/98 recorded 0% sensitivity (only one epidemic occurred in that year, but it was misclassified as a non-epidemic). The specificity, however, was very high (98%).

---

8 Validation Test. **Sensitivity**: the proportion of epidemics that the model correctly identified as epidemics. **Specificity**: the proportion of non-epidemics that the model correctly identified as non-epidemics. **Positive predictive value**: the likelihood that an epidemic predicted by the model really is an epidemic. **Negative predictive value**: the likelihood that a non-epidemic predicted really is a non-epidemic. **Accuracy**: the percentage of epidemics and non-epidemics correctly predicted. Except where otherwise noted, a cut-off probability of 0.5 was used to define an epidemic.
Table 6.16  Sensitivity, specificity, positive predictive values, and negative predictive values for the late warning model in Murray Region 2, by year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td></td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>92/93</td>
<td>100%</td>
<td></td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>93/94</td>
<td></td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>94/95</td>
<td></td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>95/96</td>
<td></td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>96/97</td>
<td>26%</td>
<td>60%</td>
<td>71%</td>
<td>18%</td>
</tr>
<tr>
<td>97/98</td>
<td>0%</td>
<td>100%</td>
<td>-</td>
<td>96%</td>
</tr>
<tr>
<td>98/99</td>
<td></td>
<td>96%</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

All years 66% 98% 91% 91%

The accuracy of the model for all years (the percentage of epidemics and non-epidemics accurately predicted) was high (Table 6.17). Accuracy decreased only marginally as the threshold for determining the probability of an epidemic increased - from 91% to 86%.

Table 6.17  Accuracy of the late warning model for Murray Region 2, using varying cut-off values to differentiate the probability of an epidemic, by year.

<table>
<thead>
<tr>
<th>Year</th>
<th>0.5 cut-off</th>
<th>0.7 cut-off</th>
<th>0.9 cut-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>92/93</td>
<td>100</td>
<td>83</td>
<td>63</td>
</tr>
<tr>
<td>93/94</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>94/95</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>95/96</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>96/97</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>97/98</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>98/99</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
</tbody>
</table>

All years 91 89 86

6.7.4 Combined region models

Data from Region 1 and Region 2 were pooled (see Chapter 5 for the method). The purpose was to examine whether the RRv bioclimatic regions were a useful method for grouping SLAs into climatically homogeneous groupings. The total number of SLAs in the combined region was 38. The total number of epidemics for the period was 70, and the number of non-epidemics was 234.
6.7.4.1 Combined regions early warning model

The combined model (with no region interactions) had eight variables:
- lagRainDays_456, RainDays_1, Rain_23, SST_2, RainDays_3, MaxTmp_5, AbsMaxTmp_5, Humid_5

Two predictor variables were common to both regions: lagged Oct-Dec rain days and July rain days. The Likelihood Ratio test indicated that July rain days had a (just) significantly different effect in the two regions ($\chi^2 = 4.06; P < 0.044$). Therefore parameters for this variable were not allowed to differ (i.e., separate coefficients were estimated for both regions). Lagged rain days (Oct-Dec) did not demonstrate a significant regional effect ($\chi^2 = 0.38; P<0.548$), and parameters were allowed to differ. The other variables that were not shared were given region interaction terms. The final logistic regression early warning model for the combined regions had the form:

$$\log \left[ \frac{P}{1-P} \right] = \alpha + \beta_1 \text{lagRainDays}_456 + \beta_2 \text{RainDays}_1 \times \text{I}_\text{REG} + \beta_3 \text{Rain}_23 \times \text{I}_\text{REG} + \beta_4 \text{SST}_2 \times \text{I}_\text{REG} + \beta_5 \text{RainDays}_3 \times \text{I}_\text{REG} + \beta_6 \text{MaxTmp}_5 \times \text{I}_\text{REG} + \beta_7 \text{AbsMaxTmp}_5 \times \text{I}_\text{REG} + \beta_8 \text{Humid}_5 \times \text{I}_\text{REG}$$

where $\text{I}_\text{REG}$ was a term to denote a region interaction.

Table 6.18 shows the variables, their coefficients, standard errors, odds ratios, and 95% confidence intervals for the combined region early warning model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>SE</th>
<th>P value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>One year lag rain days (Oct-Dec)</td>
<td>0.34</td>
<td>0.089</td>
<td>0.000</td>
<td>0.20 - 0.56</td>
</tr>
<tr>
<td>Rain days (July)* Region 1</td>
<td>2.36</td>
<td>0.739</td>
<td>0.006</td>
<td>1.27 - 4.36</td>
</tr>
<tr>
<td>Rain days (July)* Region 2</td>
<td>0.58</td>
<td>0.177</td>
<td>0.074</td>
<td>0.31 - 1.05</td>
</tr>
<tr>
<td>Rainfall (Aug-Sept)* Region 1</td>
<td>1.73</td>
<td>0.369</td>
<td>0.011</td>
<td>1.13 - 2.62</td>
</tr>
<tr>
<td>Rainfall (Aug-Sept)* Region 2</td>
<td>0.58</td>
<td>0.123</td>
<td>0.011</td>
<td>0.38 - 0.88</td>
</tr>
<tr>
<td>SST (Aug)* Region 1</td>
<td>0.92</td>
<td>1.80</td>
<td>0.964</td>
<td>0.02 - 43.57</td>
</tr>
<tr>
<td>SST (Aug)* Region 1</td>
<td>41.6</td>
<td>93.15</td>
<td>0.096</td>
<td>0.51 - 3349.41</td>
</tr>
<tr>
<td>Rain days (Sept)* Region 1</td>
<td>0.75</td>
<td>0.302</td>
<td>0.472</td>
<td>0.33 - 1.65</td>
</tr>
<tr>
<td>Rain days (Sept)* Region 2</td>
<td>1.97</td>
<td>0.912</td>
<td>0.142</td>
<td>0.79 - 4.88</td>
</tr>
<tr>
<td>Absolute max temp (Nov)* Region 1</td>
<td>0.02</td>
<td>0.028</td>
<td>0.022</td>
<td>0.00 - 0.55</td>
</tr>
<tr>
<td>Absolute max temp (Nov)* Region 2</td>
<td>70.2</td>
<td>137.37</td>
<td>0.030</td>
<td>1.51 - 3251.56</td>
</tr>
<tr>
<td>Mean max temp (Nov)* Region 1</td>
<td>1.13</td>
<td>0.806</td>
<td>0.865</td>
<td>0.27 - 4.58</td>
</tr>
<tr>
<td>Mean max temp (Nov)* Region 2</td>
<td>0.27</td>
<td>0.225</td>
<td>0.118</td>
<td>0.05 - 1.39</td>
</tr>
<tr>
<td>Relative humidity min temp (Nov)* Region 1</td>
<td>0.34</td>
<td>0.138</td>
<td>0.008</td>
<td>0.15 - 0.75</td>
</tr>
<tr>
<td>Relative humidity min temp (Nov)* Region 2</td>
<td>2.68</td>
<td>1.10</td>
<td>0.017</td>
<td>1.19 - 6.00</td>
</tr>
</tbody>
</table>

* Model $R^2$ (percentage of variance explained by model) = 81%
Model fit

The $R^2$ value (81%) suggests that the model explained a reasonable proportion of the variation in the outcome. The Hosmer and Lemeshow goodness-of-fit test was 0.43 (prob $>\text{chi}^2 = 0.99$), which suggests that the model fit the data well. Several variables had wide confidence intervals, which indicates that these coefficients were not reliably estimated. For the unvalidated model, 91% of observed values were correctly predicted.

Cross-validation

The results of the validation test are provided in Table 6.19. The overall sensitivity and positive predictive values were low - 51% and 53% respectively. The specificity and negative predictive values were both reasonable (86%). 72% of epidemics that occurred in the regions in 1992/93 were predicted, but only 23% of epidemics in 1996/97 were predicted.

**Table 6.19** Sensitivity, specificity, positive predictive values, and negative predictive values for the combined regions early warning model, all years.

<table>
<thead>
<tr>
<th>Year</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>100%</td>
<td>89%</td>
<td>20%</td>
<td>100%</td>
</tr>
<tr>
<td>92/93</td>
<td>72%</td>
<td>100%</td>
<td>100%</td>
<td>17%</td>
</tr>
<tr>
<td>93/94</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>94/95</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>95/96</td>
<td>-</td>
<td>92%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>96/97</td>
<td>23%</td>
<td>88%</td>
<td>88%</td>
<td>23%</td>
</tr>
<tr>
<td>97/98</td>
<td>100%</td>
<td>38%</td>
<td>4%</td>
<td>100%</td>
</tr>
<tr>
<td>98/99</td>
<td>50%</td>
<td>97%</td>
<td>50%</td>
<td>97%</td>
</tr>
<tr>
<td>All years</td>
<td>51%</td>
<td>86%</td>
<td>53%</td>
<td>86%</td>
</tr>
</tbody>
</table>

9 **Validation Test. Sensitivity:** the proportion of epidemics that the model correctly identified as epidemics. **Specificity:** the proportion of non-epidemics that the model correctly identified as non-epidemics. **Positive predictive value:** the likelihood that an epidemic predicted by the model really is an epidemic. **Negative predictive value:** the likelihood that a non-epidemic predicted really is a non-epidemic. **Accuracy:** the percentage of epidemics and non-epidemics correctly predicted. Except where otherwise noted, a cut-off probability of 0.5 was used to define an epidemic.
6.7.4.2 Combined regions late warning model

The combined model (with no region interactions) had eleven variables:

- $R_{D1AY456L}$, $R_{D1AY1}$, $R_{NTOT23}$, $R_{D1AY3}$, $ABS_{MAX5}$, $T_{MAX5}$, $ABS_{MAX6}$, $VP_{78}$, $RH_{MAX8}$, $ABS_{MIN8}$, $T_{MAX8}$

Both regions shared only one. The Likelihood Ratio test indicated that only one predictor variable (lagged Oct-Dec rain days) did not have a different effect in the two regions ($\chi^2 = 2.09; \ p < 0.148$), and parameters for this variable were allowed to differ. The other variables in the model were given region interaction terms. The final logistic regression late warning model for the combined regions had the form:

$$\log \left( \frac{P}{1-P} \right) = a + \beta_1 (\text{lagRainDays}_{456}) + \beta_2 (\text{RainDays}_{1xIREc}) + \beta_3 (\text{Rain}_{23xIREc}) + \beta_4 (\text{Rain}_{3xIREc}) + \beta_5 (\text{RainDays}_{3xIREc}) + \beta_6 (\text{Rain}_{6xIREc}) + \beta_7 (\text{Rain}_{8xIREc}) + \beta_8 (\text{MaxTmp}_{7xIREc}) + \beta_9 (\text{AbsMaxTmp}_{1xIREc}) + \beta_{10} (\text{MinMax}_{1xIREc})$$

Table 6.20 shows the variables, their coefficients, standard errors, odds ratios, and 95% confidence intervals for the combined regions late warning model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeffs</th>
<th>SE</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>One year lag rain days (Oct-Dec)</td>
<td>0.39</td>
<td>0.14</td>
<td>0.01</td>
<td>0.19-0.79</td>
</tr>
<tr>
<td>Rain days (July)* Region 1</td>
<td>0.82</td>
<td>0.35</td>
<td>0.64</td>
<td>0.36-1.87</td>
</tr>
<tr>
<td>Rain days (July)* Region 2</td>
<td>1.49</td>
<td>0.70</td>
<td>0.39</td>
<td>0.59-3.73</td>
</tr>
<tr>
<td>Rainfall (Aug-Sept)* Region 1</td>
<td>1.56</td>
<td>0.33</td>
<td>0.39</td>
<td>1.02-2.36</td>
</tr>
<tr>
<td>Rainfall (Aug-Sept)* Region 2</td>
<td>0.67</td>
<td>0.14</td>
<td>0.06</td>
<td>0.43-1.01</td>
</tr>
<tr>
<td>Rainfall (Sept)* Region 1</td>
<td>1.18</td>
<td>0.51</td>
<td>0.71</td>
<td>0.50-2.74</td>
</tr>
<tr>
<td>Rainfall (Sept)* Region 2</td>
<td>1.16</td>
<td>0.61</td>
<td>0.78</td>
<td>0.41-3.25</td>
</tr>
<tr>
<td>Absolute max temp (Nov)* Region 1</td>
<td>0.33</td>
<td>0.25</td>
<td>0.14</td>
<td>0.07-1.42</td>
</tr>
<tr>
<td>Absolute max temp (Nov)* Region 2</td>
<td>7.16</td>
<td>6.41</td>
<td>0.13</td>
<td>1.23-41.34</td>
</tr>
<tr>
<td>Mean max temp (Nov)* Region 1</td>
<td>0.11</td>
<td>0.18</td>
<td>0.17</td>
<td>0.00-2.59</td>
</tr>
<tr>
<td>Mean max temp (Nov)* Region 2</td>
<td>0.69</td>
<td>1.29</td>
<td>0.84</td>
<td>0.01-26.92</td>
</tr>
<tr>
<td>Absolute max temp (Dec)* Region 1</td>
<td>6.35</td>
<td>6.67</td>
<td>0.08</td>
<td>0.80-49.83</td>
</tr>
<tr>
<td>Absolute max temp (Dec)* Region 2</td>
<td>0.05</td>
<td>0.06</td>
<td>0.01</td>
<td>0.00-0.52</td>
</tr>
<tr>
<td>Vapour pressure (Jan-Feb)* Region 1</td>
<td>0.14</td>
<td>0.23</td>
<td>0.22</td>
<td>0.00-3.21</td>
</tr>
<tr>
<td>Vapour pressure (Jan-Feb)* Region 2</td>
<td>181.28</td>
<td>395.05</td>
<td>0.02</td>
<td>2.53-12980.49</td>
</tr>
<tr>
<td>Relative humidity max temp (Feb)* Region 1</td>
<td>0.41</td>
<td>0.35</td>
<td>0.30</td>
<td>0.07-2.21</td>
</tr>
<tr>
<td>Relative humidity max temp (Feb)* Region 2</td>
<td>1.54</td>
<td>1.44</td>
<td>0.64</td>
<td>0.24-9.60</td>
</tr>
<tr>
<td>Absolute min temp (Feb)* Region 1</td>
<td>8.20</td>
<td>10.07</td>
<td>0.09</td>
<td>0.73-91.14</td>
</tr>
<tr>
<td>Absolute min temp (Feb)* Region 2</td>
<td>0.14</td>
<td>0.18</td>
<td>0.12</td>
<td>0.01-1.69</td>
</tr>
<tr>
<td>Mean max temp (Feb)* Region 1</td>
<td>4.25</td>
<td>6.35</td>
<td>0.33</td>
<td>0.22-79.71</td>
</tr>
<tr>
<td>Mean max temp (Feb)* Region 2</td>
<td>0.35</td>
<td>0.59</td>
<td>0.53</td>
<td>0.01-9.34</td>
</tr>
</tbody>
</table>

* Model $R^2$ (percentage of variance explained by model) = 83%
Model fit

The $R^2$ value (83%) suggests that the model explained a reasonable proportion of the variation in the outcome. The Hosmer and Lemeshow goodness-of-fit test was 0.87 (prob $> \chi^2 = 0.99$), which suggests that the model fit the data well. The coefficients for several variables, however, were poorly estimated. For the unvalidated model, 93% of observed values were correctly fitted.

Cross-validation

The results of the validation test\(^\text{10}\) are provided in Table 6.21. The overall sensitivity was extremely low (34%). The specificity and negative predictive values were high (93% and 83% respectively). Only 31% of the epidemics that occurred in 1992/93 were predicted, and only 37% of the epidemics in 1996/97 were predicted.

<table>
<thead>
<tr>
<th>Year</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>0%</td>
<td>86%</td>
<td>0%</td>
<td>97%</td>
</tr>
<tr>
<td>92/93</td>
<td>31%</td>
<td>100%</td>
<td>100%</td>
<td>7%</td>
</tr>
<tr>
<td>93/94</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>94/95</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>95/96</td>
<td>-</td>
<td>100%</td>
<td>-</td>
<td>100%</td>
</tr>
<tr>
<td>96/97</td>
<td>37%</td>
<td>63%</td>
<td>79%</td>
<td>21%</td>
</tr>
<tr>
<td>97/98</td>
<td>0%</td>
<td>100%</td>
<td>-</td>
<td>97%</td>
</tr>
<tr>
<td>98/99</td>
<td>100%</td>
<td>78%</td>
<td>20%</td>
<td>100%</td>
</tr>
<tr>
<td>All years</td>
<td>34%</td>
<td>93%</td>
<td>60%</td>
<td>83%</td>
</tr>
</tbody>
</table>

\(^{10}\) Validation Test. Sensitivity: the proportion of epidemics that the model correctly identified as epidemics. Specificity: the proportion of non-epidemics that the model correctly identified as non-epidemics. Positive predictive value: the likelihood that an epidemic predicted by the model really is an epidemic. Negative predictive value: the likelihood that a non-epidemic predicted really is a non-epidemic. Accuracy: the percentage of epidemics and non-epidemics correctly predicted. Except where otherwise noted, a cut-off probability of 0.5 was used to define an epidemic.
6.7.5 Summary

In relation to the hypotheses outlined in Section 6.5, the following were found:

1. Low late spring to summer rainfall in the previous year (late winter/spring) predicted epidemics in both regions, and had a common effect in all the models.

2. High rainfall (late winter to early spring) in the current year predicted epidemics. Total rainfall and the number of rain days for the months of July, August, and September predicted epidemics in the early and late warning models for both regions.

3. The SOI, although significantly associated with epidemics in single variable analysis, was highly correlated with the rainfall variables. It was not a predictor in any of the models. The same was generally true for SST, although August SST was a variable in the early warning model in Region 2.

4. Low maximum temperatures in late spring and early summer predicted epidemics. Mean maximum, or absolute maximum, temperatures in November were negative predictors in all the models. Absolute maximum temperature in December was a negative predictor in the late warning model in Region 2.

5. Summer rainfall – although positively associated with epidemics in single variable analysis – did not predict epidemics in either of the late warning models.

6. High late summer temperatures (for February) predicted epidemics. Absolute minimum temperature was a positive predictor in Region 1, and mean maximum temperature was a positive predictor in Region 2. There was no evidence that extreme temperatures negatively affected the development of epidemics in the study period.

Overall, the accuracy of epidemic and non-epidemic prediction was good to excellent for both the early and late warning models in both regions (Table 6.22 provides a summary).
Table 6.22 Accuracy* of the early warning and late warning models for the two Murray study regions, by year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Early Warning</th>
<th>Late Warning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Region 1</td>
<td>Region 2</td>
</tr>
<tr>
<td>1991/92</td>
<td>71</td>
<td>100</td>
</tr>
<tr>
<td>1992/93</td>
<td>64</td>
<td>79</td>
</tr>
<tr>
<td>1993/94</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>1994/95</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>1995/96</td>
<td>100</td>
<td>88</td>
</tr>
<tr>
<td>1996/97</td>
<td>64</td>
<td>63</td>
</tr>
<tr>
<td>1997/98</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>1998/99</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>All years</td>
<td>88</td>
<td>79</td>
</tr>
</tbody>
</table>

* Accuracy = percentage of epidemics and non-epidemics correctly predicted in each year. A cut-off probability of 0.5 was used to define an epidemic for this table.

All the models predicted non-epidemics better than epidemics (Table 6.23 is a summary). In Region 1, the early warning model had a sensitivity of only 62% (16 of 26 epidemics). Prediction improved dramatically for the late warning model, which had a sensitivity of 96% (25 of 26 epidemics). In Region 2, the early warning model had a sensitivity of 73% (32 of 44 epidemics), but the late warning model had a lower sensitivity of 66% (29 of 44 epidemics).

Table 6.23 Sensitivity and specificity of the early warning and late warning models for the two Murray study regions, all years.

<table>
<thead>
<tr>
<th>Early Warning</th>
<th>Late Warning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1</td>
<td>Predicted</td>
</tr>
<tr>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>62%</td>
</tr>
<tr>
<td>Specificity</td>
<td>95%</td>
</tr>
<tr>
<td>Region 2</td>
<td>Predicted</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>73%</td>
</tr>
<tr>
<td>Specificity</td>
<td>81%</td>
</tr>
</tbody>
</table>
The combined regions models had substantially reduced sensitivity than either of the separate region models. In the combined early warning model, 51% of epidemics (36 of 70 epidemics) were predicted. In the late warning model, only 34% were predicted (24 of 70 epidemics). This confirms the importance of taking into account regional differences when building predictive models for RRv disease.

6.8 Discussion

I developed parsimonious early and late warning models for two regions in the Murray area, using climatic factors as explanatory variables. The sensitivity and specificity values for the models indicate that climate data alone can be used to predict the probability of epidemics with moderate to very good results. The advance notice provided by the timing of the models' predictions (early December for the early warning models, and early March for the late warning models) is such that there is enough time for response to be conducted to control mosquito population growth and reduce the risk of transmission to humans. The means by which these models may be used to enhance public health planning and response for RRv disease is discussed in Chapter 8.

6.8.1 Limitations of the study

The following factors must be considered in the evaluation of the results.

- The principal limitation of this study lies with the notification data, which are reported by place of residence rather than place of suspected infection. I assumed that in rural areas with relatively large SLAs, work and recreation (and hence transmission) generally occur within the same SLA. For a small proportion of cases this assumption may not be true. This would result in a mis-classification of these cases, and weaken the association between the exposure variables and the outcome.
- The relatively low sensitivity of two of the models (62% and 66%) suggests that factors other than climate may be involved in determining the timing of epidemics, or that the temporal unit of analysis was too coarse to capture intra-monthly climate fluctuations (particularly in temperature). Simple predictive models such as these ones could form the basis for cost-effective allocation of resources to improve data collection (e.g. mosquito trapping). Additional data, such as the number of mosquitoes or the level of virus activity, may improve the accuracy of the models.
Although high mosquito numbers do not necessarily indicate that an epidemic will occur, they are generally a precondition for one. In the two Murray Valley regions these data were not routinely available.

- The analyses were conducted with constraints on the disease data time series. The likelihood of confounders explaining the observed relationship will be greater for a short time series. In a study of eight years, it is possible that confounding factors may coincide with the ENSO cycle (Kovats et al. 1999). There were an unusual number of El Niño events that occurred during the study period (from 1990/91-1994/95, and again in 1997/98: Climate Prediction Center 2001), and I cannot be definite about the strength of the predictive power of SST in the Region 2 early warning model. One La Niña (cool phase) occurred in 1998/99 (weak in winter, and moderate in spring/early summer).

- Both the model and the validation method are subject to error. The cross-validation method assessed how well the variables in each of the models were able to predict epidemics, rather than the predictive performance of the estimated coefficients for the eight-year model. In other words, the results of the validations reflect the relationship between the predictions and the observations, and do not give a precise measure of how well the models perform in absolute terms. A “gold standard” method would have taken the model (with its estimated coefficients) and applied it to the same region for a different period. However, additional disease data were not available. Epidemics occurred infrequently in this data set (in general only in two of the eight years). In the rotating validation process, when one year of epidemic information was removed only half the data on epidemics remained to derive the parameter estimates of the prediction model for that year. Given there are several variables in each model, it may be that some of these predicted epidemics by chance.

- Multiple comparison is an intrinsic problem with the exploratory analysis approach that was used to select the candidate variables for the multivariable modelling. I attempted to manage this by only considering those variables that had been associated with the outcome via a biologically plausible mechanism.

- The fact that the majority of epidemics in the period occurred during two of the eight years challenges the statistical assumption of independence that underlies logistic regression modelling. RRv disease epidemics will in some cases have dependent association, by virtue of the mode of infection. To manage possible dependence between units would have required a computer intensive approach. The amount of time needed for such analysis would not have been feasible for this study. Instead, I used logistic regression as a tool for more rapid analysis. Consequently, the estimate
of effect is likely to be unbiased, but the confidence intervals generated cannot be interpreted too literally.

- In the Murray Region, mosquito control interventions were minimal throughout the period. However, estimates of the timing and effectiveness of mosquito sprays, and of variation in irrigation practice by SLA, would strengthen the findings. It is proposed that data on the extent of irrigation coverage in Australia will become available over the next several years, as part of the National Heritage Trust data collection. Such data would be useful for future similar work.

With the above comments in mind, the following points can be made about the results.

### 6.8.2 Host immunity

Ideally, extrinsic (i.e., climatic) factors should be combined with host-virus population dynamics for accurate epidemic prediction (Hay et al. 2000). Although host population dynamics were not able to be directly modelled (i.e., via animal density counts or serosurvey results), the models successfully used covariate patterns - low rainfall in the preceding year combined with suitable climate in the following - to account for the influence of vertebrate host immunity levels.

A finding common to all models was that rainfall in the late spring to summer months of the lagged year (the year preceding the epidemic year) was negatively associated with epidemics. Kangaroo populations take about one year to complete the breeding cycle (Caughley et al. 1987). If rainfall is low in the late spring to summer period, this will generally be accompanied by low mosquito population numbers, and minimal cycling of the virus between vectors and hosts. Consequently, the host population immunity level will reduce. If the following year, high late spring to summer rainfall would lead to abundant mosquito activity, amplification of the virus through the susceptible hosts would occur, and humans would be more at risk of exposure. For a large outbreak to occur in epizootic regions, it is likely that several years of below average rainfall is needed to sufficiently reduce the reservoir host immunity. In the study period, there was a four year break between epidemics.

Dhileepan (1996) reported very high numbers of mosquitoes in the Murray Valley area in the summers of both 1992/93 and 1993/94. The requirement for low levels of host immunity, moreover, may explain why no epidemics were reported in 1993/94, even
though the other climatic conditions in that year were suitable (i.e., there was above average rainfall and temperatures).

6.8.3 The influence of rainfall

Late winter/early spring rainfall variables, as expected, proved to be important predictors of epidemics in both regions. Both the total amount of rainfall and the number of rain days in these seasons were significant predictors.

Summer rainfall excessively above normal has been coincident with the major RRv disease epidemics in the regions, and summer rainfall variables were expected to be positively associated with epidemics. Although there were some significant associations in the single variable analysis, essentially two different rainfall patterns were observed between the two main epidemic years, and as a consequence rainfall variables did not emerge as predictors of epidemics in either of the late warning models. In the first epidemic year (1992/93), heavy rainfall of nearly twice the long-term mean commenced in August, and above average rainfall was sustained throughout the spring into mid summer. This was one typical pattern suggested by the historical record. In the second epidemic year (1996/97), higher than average rainfall commenced even earlier (June) and continued until September in both regions. This was followed by average to below average summer rainfall in both regions.

Sustained winter and spring rainfall could enhance transmission in these regions by two mechanisms. First, it would allow for the early and prolific breeding of *Aedes* populations, and an extended period of virus build-up between the initiating vectors and the host populations. Second, prolonged heavy spring rainfall would raise the water table, reducing absorption and run-off. Even if summer rainfall was low, pools of surface water could remain throughout the early months of summer, providing sites for *Cx annulirostris* breeding.

Although high rainfalls were recorded in some individual late winter months in other years (e.g. 1991 and 1995), they were followed by lower than average rainfalls in July and August. This may have been a factor in aborting an epidemic in those years.
6.8.3.1 The role of irrigation

An anomalous irrigation pattern for the Murray River provides a possible explanation for the unusual occurrence of widespread RRv disease epidemics during a summer of average to below average rainfall (1996/97).

The Murray River and its tributaries border almost all of the SLAs in Region 1, and most in Region 2. The extent of water in an otherwise low rainfall (and variable) region undoubtedly enables limited mosquito breeding to occur in dry years (Russell 1986a). Even so, greatly improved irrigation practices have dramatically reduced the availability of water sites for mosquito breeding (Dhileepan 1996). Dhileepan has noted that irrigation changes in these regions since the mid 1980s include (i) laser grading to minimise pooling water and stagnation, and (ii) reduced flood irrigation practice due to the cost of irrigation water and education relating to the risks of arboviral disease. The extent to which irrigation influenced the mosquito breeding cycles in the study area, over and above the contribution of rainfall, could not be quantified in this study.

The Hume Reservoir modulates water levels for the Murray River, influencing flood patterns downstream. The predominant irrigation method used by industries along the Murray River is the annual release of water throughout the growing season from the Reservoir. The Reservoir is subject to an annual cycle of filling (from rainfall) and then release, which results in flooding for a restricted number of irrigation leases along the banks of the river downstream. The Reservoir is filled from June to October each year, and water is released from (typically) August. Under normal conditions, a maximum irrigation release of 25,000 mega litres/day is permitted, and flooding under this regulated flow does not involve large areas (Mel Jackson pers. comm. 2001).

In 1992, total spring rainfall from August to October was almost twice the long-term average for both study regions. Consequently, the water discharged from the Reservoir in October 1992 was higher than usual for several weeks - a reflection of the large volume of water in the Reservoir. Figure 6.21 is a time series of the Murray River flow during the study period, measured at Albury, New South Wales. Note the large peak during the spring of 1992, and an extended peak in October 1996.

In 1996, total winter and spring rainfall for June to September was also well above average, and some extra water would normally have been released from the Reservoir in
Figure 6.21  Amount of water released from the Hume Reservoir into the Murray River per week (1991 to 1999). Measured downstream at Doctor's Point, Albury, New South Wales.

Figure 6.22  Amount of water released from the Hume Reservoir into the Murray River per week (September to December 1996). Measured downstream at Doctor's Point, Albury, New South Wales.
October to reflect this. However, due to fears of a crack in the Reservoir wall a “special release” of an additional excessive volume of water was authorised in October 1996 (Mel Jackson pers. comm. 2001). This release artificially maintained the Murray River flow at “flood” levels for more than 30 days. Flow levels throughout October were persistently above 25,000 day/day (for two weeks they were more than 90,000 day/day, and did not return to normal levels for six weeks). Figure 6.22 is a detail of flow levels for the period September to December 1996.

With this volume of water, the land area under inundation increases rapidly, and billabongs and low depressions around the river fill up (River Murray Commission September 1976). The relationship between the flow measured at Albury and the area of inundated land is non-linear (see Figure 6.23). The sustained high levels of flow in October 1996 resulted in extensive and prolonged flooding of the predominantly flat country-side surrounding the Murray River, and in insurance claims for damage to agriculture (Mel Jackson pers. comm. 2001).

**Figure 6.23** Volume of water released from the Hume Reservoir and estimated land under inundation around Albury, New South Wales.

![](image)

*Data source: Murray Darling Basin Commission.*
The above average winter and spring rainfalls of 1996 are likely to have raised the water table along the Murray River, resulting in minimum absorption of the widespread flooding that occurred throughout October and remained into November and possibly December in places. So although the summer rainfall of 1996 was below average, the flood conditions across the region, coupled with the high water table sustained from the winter/spring falls, would have provided ideal conditions for (i) amplification of the virus within the natural cycle by spring breeding mosquitoes, and (ii) an abundance of Cx annulirostris during the summer period.

In summary, these findings support the hypothesis that extensive summer flooding is one of the precursors needed for an outbreak of RRv disease in these regions. This situation may be achieved via heavy spring rainfall followed by heavy summer rainfall, or by heavy spring rainfall and widespread flooding from atypical irrigation practices. It is also possible that above average summer rainfall on its own will be sufficient to support the development of summer breeding mosquito populations, although this pattern did not occur during the period and was not able to be tested. Above average spring rainfall in both regions was strongly and significantly associated with the initiation of epidemics, but it is not possible to conclude that it is an essential precursor. It may be one of several mechanisms that lead to the build-up of sufficient virus and mosquitoes in these regions.

6.8.4 The influence of temperature

The finding of a negative association between November temperatures and epidemics is consistent with the spring breeding habits of Cx australicus, the principal amplification vector, which shows greatest population growth at lower temperatures (range 6-20°C) (Dhileepan 1996). Lindsay has also observed that RRv can replicate to infectious titres in mosquitoes through a wide temperature range (Lindsay 1995). To the extent that Ae. vigilax is involved as a secondary amplification vector in this area – not entirely clear from the mosquito field studies (Dhileepan 1996, Russell 1986b) – this negative association is also consistent with a recent study of temperature influence on the RRv vector competence of this species. In laboratory experiments, Kay and Jennings (2002) observed that moderate temperatures (18 to 25°C) were optimal for RRv transmission in preference to higher temperatures (such as 32°C).
February temperature variables in both regions were positive predictors of epidemics. Several studies have identified the increased survival of infected female mosquitoes as the single most important factor in contributing to infections (de Moor and Steffens 1970, Scott et al. 1983). At higher temperatures vectorial capacity is increased, due to a reduction in the extrinsic incubation period (Weinstein 1997), and the mosquito generation time is shortened. The life span of \textit{Cx annulirostris} is longest when mean ambient temperatures persist at around 25-27°C (Mottram et al. 1986, Dhileepan 1996), and “summer temperatures in the upper range” have been observed to favour large population growth in \textit{Cx annulirostris} (Dhileepan 1996).

In Region 1, which abuts the Australian inland desert and is affected by temperature extremes, February minimum temperatures were positively associated with epidemics. Mosquito reproduction can be interrupted or terminated by even several days of cold weather in the summer months (Hawkes et al. 1985). A study in the RRv disease endemic region of temperate northern Australia also found increases in minimum temperature to be positively associated with case incidence (Tong et al. 1998).

The risk of an epidemic year was three times higher (OR=2.94, CI=1.7-5.1, \( P<0.002 \)) when absolute minimum temperatures in August were above zero (the early spring period, when \textit{Aedes} species are breeding in this region). Temperatures above freezing are more conducive to the early breeding of \textit{Aedes} populations. These species act as initial amplifiers of the virus between vector and vertebrate host populations.

\subsection*{6.8.5 Analysis within RRv bioclimatic regions}

Several comments are suggested about the value of using RRv bioclimatic regions to define the boundaries for spatial analyses in a project such as this.

First, combining the two Murray study regions resulted in a lower predictive sensitivity than either of the separate models could provide on their own (the combined early and late warning models predicted only 51\% and 34\% of epidemics respectively). This suggests that working at a larger scale – at least in this RRv epidemic part of Australia – would not improve prediction accuracy.
Second, the heterogeneity of the explanatory variables in each of the regions meant that region-specific parameters had to be attached to each variable in the combined models. The early warning combined model had fifteen variables (compared to five each for the Region 1 and 2 early warning models). The late warning combined model had 21 variables (compared to six each for the Region 1 and 2 models). Therefore, the combined models were far less parsimonious than either of the separate study region models.

This approach tested whether epidemics could be accurately predicted at the regional scale in which public health planning and policy occur. This does not mean that monitoring of local incidence of RRv disease and meteorological records at a finer resolution (such as a specific town or community) would be less predictive. For several important reasons discussed more fully in Chapter 8, most particularly the resource inefficiencies of working at this scale, such an approach is not appropriate for most public health regions in Australia.

6.8.6 Precision and validity of the models

The models (early and late) developed for both regions were parsimonious (with five or six variables in each), which simplifies the process of collecting and analysing data. The models had varying levels of validity, ranging from moderate to very high. Both regions had at least one model with a sensitivity greater than 70% (96% for the early warning model in Region 2, and 73% for the late warning model in Region 1). In terms of response, vector control activities in these regions are best conducted in early December. The model sensitivity of 62% in Region 1 (or about 6 chances out of 10 of an epidemic occurring) is probably not a good enough indicator on its own as a basis for public health action. But it could be used as an indicator of the need for further data collection (e.g. surveillance, mosquito trapping, etc. – discussed in Chapter 8).

6.8.7 Implications for climate change

Summers in south-west New South Wales are expected to become 15% wetter by 2030 (20% by 2070) and springs will become about 10% drier (35% by 2070). The number of extremely dry springs is projected to more than double after 2020, as does the number of extremely wet summers (CSIRO 2001).

The effect of these changes on the breeding and survival of mosquitoes and vertebrate host populations, even in a region as “localised” as the southeast of Australia, is difficult
to anticipate. Russell (1998b) has speculated that *Aedes* populations in dry areas (such as Region 1) may be adversely affected by the projected decrease in winter rainfall in this part of Australia, possibly delaying or precluding virus activity. Given the findings in this chapter, I would expect this to cause a reduction in the frequency of epidemics in years when the "winter rainfall pattern" might have occurred. Conversely, the predicted increase in summer rainfall may increase the availability of mosquito habitat that, combined with higher average temperatures, may lead to higher humidity, a lengthened season of abundance and greater transmission levels (Russell 1998b).

Although the SOI and SST were highly predictive of epidemics in single variable analyses, in general both were highly correlated with (and better predicted by) other precipitation variables. The exception was August SST, a predictor in the Region 2 early warning model. It is not surprising that SST was a better predictor than the SOI of rainfall excess (and hence epidemics) in this instance. The calculation of the SOI incorporates air pressure values from both sides of the Pacific basin (Darwin and Tahiti), whereas SST records the rise and fall of temperatures directly in the "El Niño" region of the eastern Pacific. Therefore, SST is likely to be a more specific indicator of changes in Pacific Ocean temperatures (Janette Lindesay pers. comm. 2000).

### 6.8.8 Generalisability of results

These findings about the timing and direction of association of climatic factors on vector breeding and survival, and on host immunity, are likely to be relevant to adjacent areas (such as other SLAs in the semi-arid and temperate regions surrounding the Murray River).

The combined regions analysis suggests that, for the type of modelling conducted in this study, generalisability of specific findings outside these RRv bioclimatic regions may not be possible.
Chapter 7

The Southwest study area: using climate and mosquito data to predict epidemics

7.1 ABSTRACT

This chapter examines the affect of climate and environment on the occurrence of epidemics of RRv disease in a temperate region of south-west Western Australia. In addition, the contribution of mosquito data to improving the prediction of epidemics was tested over and above that of climate and environment data. For the whole region, climate and environment data on their own were able to predict epidemics with moderate skill (64%) from early in December. A low number of rain days in the spring to early summer of the pre-epidemic year, combined with high sea surface temperatures and rainfall for November, and high tide heights for October-November, were the best predictors of epidemics. Mosquito surveillance data contributed significantly to the sensitivity of the model. Mosquito density data for November of the pre-epidemic year increased the sensitivity of the early warning model by 26%, and of the late warning model by 21%.
7.2 PROFILE OF THE STUDY AREA

7.2.1 Population and land use

The "Southwest" study area in south-western Western Australia (Figure 7.1) falls within the bioclimatic region that extends from Perth in the north down to Albany in the south. As per the method outlined in Chapter 5, a number of Statistical Local Areas (SLAs) in the bioclimatic region were excluded from analysis:

(i) Twenty six SLAs attached to the Perth metropolitan Statistical Division were excluded.

(ii) Four SLAs inland to the south that recorded less than 10 cases of RRv disease during the period were excluded.

Figure 7.2 shows the unedited bioclimatic region (pale shading), and the Southwest study area within it (darker shading). Figure 7.3 is a detailed map of the Southwest, with SLAs and major towns shown.

Table 7.1 Characteristics of the Southwest study area.

<table>
<thead>
<tr>
<th>Southwest region</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of SLAs</td>
<td>14</td>
</tr>
<tr>
<td>Population*</td>
<td>221 833</td>
</tr>
<tr>
<td>Area (km²)</td>
<td>22 691</td>
</tr>
<tr>
<td>Population density (per km²)</td>
<td>9.7</td>
</tr>
</tbody>
</table>

The Southwest comprises 14 SLAs (total area 22 691km²) (Table 7.1), and had a total population of 221 833 in 1996. The principal land use was rural and semi-rural, with a range of small to large towns. Tourism, forestry, agriculture (fruit and vegetable production, and viticulture), and dairy were the major industries of the region (Australian Bureau of Statistics 1998b).

7.2.2 Climate and environment

The south-western region of Australia has limited topographical relief, and is largely occupied by the Yilgarn Block, an inland plateau. The Darling Scarp runs along the western boundary of the Yilgarn Block, and separates the west coastal region from the inland plateau (Mockrin 2002). A rapid increase in rainfall can be measured from the coastal plain to the top of this escarpment (which reaches elevations of 300 to 400m), followed by a marked decrease in rainfall to the east.
Figure 7.1  Australia, showing the location of the Southwest study area (shaded).
Figure 7.2  The Southwest RRv bioclimatic region (pale shading) and the Southwest study area within it (dark shading). Major towns in the study area are highlighted.
Figure 7.3  Statistical Local Areas within the Southwest study area.
The bioclimatic region in which the Southwest study area is located is classified as having dominant winter rainfall (Australian Bureau of Meteorology 2000a). High tides inundate large areas of salt-marsh and provide ideal mosquito breeding sites for *Ae. camptorynchus* and *Ae. vigilax* (Lindsay *et al.* 1989). The coastal region from the town of Mandurah down to Busselton (the “Swan Coastal Plain”) includes several major estuaries and inlets, brackish wetlands, and freshwater lakes that are affected by tidal fluctuations. Daily tidal fluctuations off the coast are usually less than 100 cm, ranging to a mean maximum of 150 cm (Lindsay 1995). A large number of people live close to these major lakes and estuaries, and residents of Perth are most likely to holiday in these coastal areas (Lindsay 1995).

The Southwest study area is classified as temperate. On average across the area, 950 mm of rainfall is recorded a year.\(^\text{11}\) Rainfall is strongly seasonal with a pronounced summer drought (Gioia and Pigott 2000), and 88% of rain falls between the months of April to October. Winter rainfall is estimated to be of very low to low-to-moderate variability (Australian Bureau of Meteorology 2002a).\(^\text{12}\) Rainfall variability during summer (Dec-Feb) is rated as very high to extreme in the area. Figure 7.4 shows long-term average rainfall, and maximum and minimum temperature values for the Southwest.

Rainfall over south-western Australia has been weakly related to the El Niño, with an estimated 10% decrease in rain overall during such an event (Partridge 1994). La Niña events, however, have been more strongly related. In general, an increase of more than 30% in rainfall would be expected across the study area in a La Niña year.

Summers are warm to hot (mean minimum temperatures are 15°C across the study area, and mean maximum temperatures are 27°C). Winters are cold to mild (mean minimum temperatures are 8°C and mean maximum temperatures are 17°C), with infrequent frosts.

The vegetation of the study area is characterised by heath, scrub, and low woodlands along the coastal strip. Jarrah (*Eucalyptus marginata*) and marri (*Eucalyptus calophylla*) forests dominate through much of this ecoregion, with jarrah the dominant species on the Darling Scarp. Tall blackbutt (*Eucalyptus patens*), jarrah, and marri forest can all be found in river valleys. Further inland, the climate becomes drier and the study area borders on the wheat belt region of Western Australia. Vegetation here consists of open eucalypt

\(^{11}\) Based on long-term average values from 1955 to 1996: data provided by Silo (Queensland Department of Natural Resources 2000).

\(^{12}\) Calculated by the method \([((90^{\text{th}} \text{percentile})-10^{\text{th}} \text{percentile})/50^{\text{th}} \text{percentile}]\) (Australian Bureau of Meteorology 2002a).
woodlands containing wheat belt wandoo (*Eucalyptus capillosa*) and powderbark (*Eucalyptus accedens*).

**Figure 7.4** Long-term average rainfall and maximum and minimum temperature profiles for the Southwest study region. Average maximum temperature (°C) △; average minimum temperature (°C) ■; rainfall (mm) ●.

<table>
<thead>
<tr>
<th>Temp (°C)</th>
<th>Rain (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>200</td>
</tr>
<tr>
<td>25</td>
<td>160</td>
</tr>
<tr>
<td>20</td>
<td>120</td>
</tr>
<tr>
<td>15</td>
<td>80</td>
</tr>
<tr>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

7.3 **ECOLOGY OF RRv IN THE SOUTHWEST**

7.3.1 **Vectors**

The most substantial work to date on the ecology and epidemiology of RRv in Western Australia has been conducted by Dr Michael Lindsay (see 1989, 1992, 1993b, 1993c, 1996a, 1997). For the Southwest study area overall, *Ae. camptorynchus* appears to be the dominant RRv vector (Lindsay *et al.* 1989, Russell 1994). Of some 340,000 mosquitoes trapped between 1987 and 1989 in the Mandurah to Bunbury region, 85% were *Ae. camptorynchus* and 9% were *Ae. vigilax* (both salt-water breeders) (Lindsay *et al.* 1989, and see Table 7.2). *Ae. camptorynchus* is reported to be most common in autumn to late spring and *Ae. vigilax* most common in summer (Lindsay *et al.* 1989, and also Russell 1994). It has also been suggested that *An. annulipes* may play an important role in the over-wintering and initiation of RRv in the Southwest (Lindsay *et al.* 1989), although this remains speculative.
Table 7.2  Breakdown of mosquito type from trapping conducted in the Mandurah to Bunbury regions of the Southwest study area, 1988/89.

<table>
<thead>
<tr>
<th>Species</th>
<th>% of total mosquitoes caught</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ae. camptorynchus</td>
<td>85.5</td>
</tr>
<tr>
<td>Ae. vigilax</td>
<td>9.0</td>
</tr>
<tr>
<td>An. annulipes</td>
<td>0.6</td>
</tr>
<tr>
<td>Coquillettidia linealis</td>
<td>0.15</td>
</tr>
<tr>
<td>Cx annulirostris</td>
<td>0.1</td>
</tr>
<tr>
<td>Other minor species</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Adapted from (Lindsay et al. 1989).
* Previously implicated as vectors of RRv.

In most years Ae. camptorynchus is replaced by Ae. vigilax in the summer months. However, in the 1988/89 (Lindsay et al. 1989) and 1991/92 (Russell 1994) outbreaks, Ae. camptorynchus persisted longer than usual - well into summer. This provided an overlap with Ae. vigilax, and was credited as being a significant factor in allowing the ongoing transmission of virus between mosquito species in those years, and hence a large outbreak of disease in humans (Lindsay et al. 1989).

An alternate explanation is suggested by results of research into the residential proximity to waterways as a risk factor for RRv disease in the Mandurah region (reported in Lindsay et al. 1998). Of 162 RRv disease cases who lived in the Mandurah region, 64% lived within one kilometre and 99% within three kilometres of waterways. The two principal mosquito vectors in that region were Ae. camptorynchus and Ae. vigilax. Ae. camptorynchus is believed to disperse between two and five kilometres away from breeding sites. Conversely, Ae. vigilax had previously been observed to disperse up to 100 km away from breeding sites (reported in Lindsay et al. 1998). If the latter is true (and no dispersal studies of Ae. vigilax had been conducted in the Mandurah region to verify this), it may be that Ae. vigilax plays a less important role than previously thought in RRv transmission in the south-west. If it were a significant vector, cases from further afield would have been expected. Perhaps climatic and environmental conditions that enable Ae. camptorynchus to persist into early summer are sufficient pre-conditions for an outbreak.

Regular monitoring of adult mosquito populations and RRv activity has occurred since 1987 at up to 40 sites between Rockingham (north) and Dunsborough (south in the
Chapter 7 - The Southwest study area

Busselton SLA (Lindsay et al. 1996c). Priority has been given to placing traps in regions where RRv and BFv appear to be endemic, where high attack rates are reported, and where viral activity appears to commence and then spread outwards to other areas (Lindsay et al. 1998). Mosquito collections have been conducted each fortnight during late spring, summer and early autumn, and once a month during the remainder of the year.

7.3.2 Vertebrate hosts

Stanley and Choo (1966) found no evidence of RRv infection among a group of 188 domestic fowl from various regions of Western Australia. Lindsay (1995) noted that studies in the 1970s also found a very small percentage of birds and cattle showed evidence of infection with RRv, and concluded that birds and cattle were unlikely to be major hosts of RRv in Western Australia. Lindsay (1996c) made reference to the presence of large numbers of an assumed host, the Western Grey kangaroo (Macropus fuliginosis), in proximity to the wetland areas of Busselton during the large 1995/96 outbreak. Other than these references, no other published work has referred to the contribution of potential RRv hosts in the Southwest. For the purpose of testing hypotheses about host immunity, I have assumed that the Grey kangaroo, as found elsewhere in Australia, was likely to be the primary vertebrate host in the study area.

7.4 EPIDEMIOLOGY OF RRv DISEASE

7.4.1 Historical

The first documented outbreak of RRv disease in Western Australia was in 1956 (reported in Marshall and Miles 1984). Since then several large outbreaks have been reported in 1988/89 (Lindsay et al. 1989), 1991/92 (Lindsay et al. 1992), and 1995/96 (Lindsay et al. 1996c). The first continuous records of cases of epidemic polyarthritis commenced in 1975, and RRv disease became notifiable in Western Australia in 1985 (Lindsay et al. 1989). An analysis of cases from 1975 to 1988 13 (Lindsay et al. 1989) indicated that the majority of cases (60%) were reported in the temperate south-west of the State, with the rest occurring in the arid and tropical north. Most cases occurred in summer and early autumn close to the coast (within 30 km), and people between the ages of 31 and 40 years were most commonly infected.

13 Total case numbers for the period were not reported.
In the south of the State, the Swan Coastal Plain has historically recorded the largest number of cases. Several reasons for this have been suggested, including the tidal salt-marsh and wetland nature of the habitat that is ideal for mosquito breeding, and the extensive human settlements in close proximity to these sites (Lindsay 1989). A large outbreak in 1988/89 (more than 620 cases) commenced in the Swan Coastal Plain in late October, with cases spreading to other parts of the south-west of the State. The pattern of cases suggested a gradual dispersal of virus activity from the Swan Coastal Plain to other areas (Lindsay et al. 1989). Lindsay suggested several environmental and climatic factors that may have contributed to this outbreak: (i) higher than average annual mean sea level and summer daily tidal heights off the Mandurah to Bunbury coast, (ii) related to this, the SOI registered positive values (associated with a La Niña event, which brings warmer than average water to the Australian region) for 13 consecutive months from July 1988 to July 1989, and (iii) above average late spring (Oct-Nov) rainfall occurred in 1988.

In 1991/92, some 555 cases were reported in the south-west area (Russell 1994), also in the Swan Coastal Plains region. Record rainfall was recorded between November and February, although mean sea level and tide heights were normal.

In 1995/96, some 545 cases were reported between November and February, with the same coastal region most affected (Lindsay et al. 1996c). Nearly 65% of cases recorded the date of onset as January. Climatic conditions observed were: high daily rainfall for October, above average rainfall in December, high daily tide heights in late December, and above average October and November temperatures. Lindsay (1996c) has also observed that “...mild spring and summer temperature conditions were predisposing factors during previous outbreaks”, and cited two previous outbreaks with similar conditions (discussed in Lindsay et al. 1989, Lindsay et al. 1992, Russell 1994).14

---

14 I could not find a reference in these two papers to temperature in this context, so the importance of “mild” temperatures on mosquito breeding in this region remains unclear.
Table 7.3  The seasonal pattern of historic RRv disease epidemics and rainfall and tides in the Southwest study area, summarised from the literature.

<table>
<thead>
<tr>
<th>Year</th>
<th>J</th>
<th>A</th>
<th>S</th>
<th>O</th>
<th>N</th>
<th>D</th>
<th>J</th>
<th>F</th>
<th>M</th>
<th>A</th>
<th>M</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988/89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991/92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995/96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- approximate duration of RRv disease outbreaks
- months when above average rainfall was noted
- months when higher than average tide heights were recorded

To summarise (Table 7.3), two large outbreaks in the Southwest have been preceded by above average rainfalls and accompanied by high tides, both of which contributed to the flooding of coastal breeding sites. High tides were not observed prior to or during the 1991/92 epidemics, but very high rainfalls in the late spring and summer months were (conducive to an extended breeding season for *Ae. camptorynchus*). Sufficient ground water for breeding during the otherwise dry summer months is necessary to support large populations of mosquitoes. From the above observations, it appears this can be achieved via at least two mechanisms:

1. Unseasonal heavy rainfall and high tides or both during the summer months.
2. Above average mean sea level from winter onwards, accompanied by heavy spring rainfall, to sustain a sufficiently high coastal water table.

The contribution of temperatures is not clear from this history: both above average and mild spring and summer temperatures have been noted.

7.4.2 Study period

For the Southwest as a whole, 1507 notifications were recorded from July 1991 to June 1999 (Figure 7.5). There were more male notifications than female ones (ratio 1 : 0.87), whereas the national sex ratio was 1:1.03. The majority of notifications were in the 25-59 year age group (75%), with only 4% of notifications recorded in children aged less than 15 years (almost identical to the national notification pattern). The eight-year average annual incidence rate for the 14 Southwest SLAs was 85/100 000 (range 13-200/100 000). This was substantially higher than the national average (25/100 000).
Table 7.4 gives the number of epidemics in the study area by year. I used two different definitions to calculate an epidemic (see note under Table 7.4, and discussion in Section 7.5). In any given year a maximum of 14 epidemic or non-epidemics could have occurred (i.e., one per SLA in the study area). Both definitions resulted in a similar epidemic pattern across the Southwest. Widespread epidemics occurred for 1995/96, the year in which the largest number of notifications was recorded (655). No epidemics were recorded in the three years preceding 1995/96, and only few epidemics were recorded in other years. With Definition 1 there was a total of 22 epidemic (and 90 non-epidemic) SLAs during the 8-year period. With Definition 2 there was a total of 18 epidemic (and 86 non-epidemic) SLAs.

Table 7.4 Number of epidemic Statistical Local Areas in the Southwest study area by year, using two different epidemic definitions.*

<table>
<thead>
<tr>
<th>Year</th>
<th>Epidemic SLAs Definition 1</th>
<th>Epidemic SLAs Definition 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>92/93</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>93/94</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>94/95</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>95/96</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>96/97</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>97/98</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>98/99</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>22 (90)*</td>
<td>18 (94)*</td>
</tr>
</tbody>
</table>

* Definitions used were (1) the mean plus one standard deviation of all the cases recorded during the study period, and (2) the mean plus two standard deviations.
† Number of epidemics (non-epidemics) in the Southwest study area during the period.
Two minor differences are noted in the pattern between Definition 1 and 2. First, no epidemics were recorded for the year 1991/92 under Definition 2. Second, no SLAs recorded epidemics in consecutive years under Definition 2. In most instances those SLAs missed being defined as "epidemic" under Definition 2 by only 0-1 cases (i.e., the number of cases was right at the threshold). Therefore, it is likely that the different pattern between the epidemic definitions may not be significant.

Figure 7.6 shows the average number of notifications for the study period by the estimated month of onset of disease. Peak notifications were recorded in January and February (26% and 23% respectively), and the month in which the next highest number was recorded was December. This seasonal pattern is different to that of the Murray study area: epidemics there peaked some two months later.

The timing of epidemics varied spatially. In 1995/96, widespread epidemics were recorded in 12 of the 14 SLAs. In that year, the first cases were notified in October (in 1 of 12 SLAs), in November (in 3 of 12 SLAs), in December (in 1 of 12 SLAs), and in January (in 6 of 12 SLAs). The SLAs in which cases appeared earliest (October-November) were in the Swan Coastal Plain SLAs of Busselton, Capel, Bunbury, and Mandurah. This suggests that particular climatic and environmental factors in those regions (such as high tides) may have supported the early breeding of mosquitoes and the amplification of virus in the natural cycle, with infected mosquitoes spreading to other SLAs in later months. Epidemics in other years mainly occurred in the Mandurah (96/97, 98/99), Bunbury
(97/98), Capel (97/98), and Busselton (97/98) SLAs, and most of those also began earlier in the season (September, 1; October, 1; November, 3; December, 1). The epidemics that occurred in the two SLAs furthest to the north (Rockingham) and to the south (Denmark) of the Southwest commenced later, in January.

7.5 ANALYTIC METHODS PARTICULAR TO THE SOUTHWEST

In general, the methods used to develop the predictive models in the Southwest study area are the same as those used for the Murray study area (described in Chapter 5). Four extra methodological matters need mentioning. These are:

- the use of an additional outcome variable
- how the mosquito surveillance data were managed in analysis
- how the tidal height data were managed in analysis
- rationale for the reduced number of months of data used in the late warning models.

7.5.1 Additional outcome variable tested

The findings from the Murray regions prompted the questions: (i) was the epidemic definition used in those regions the most appropriate for this type of study?, and (ii) would a more specific definition result in a more reliable model? In this chapter, two different outcome variables - “epidemic (definition 1)” and “epidemic (definition 2)” - were used in the logistic regression modelling. The epidemic (definition 1) was: the number of cases in any one year that exceeded the mean plus one standard deviation of all the cases recorded during the study period (this was used for the Murray study area models). The epidemic (definition 2) was: the number of cases in any one year that exceeded the mean plus two standard deviations of all the cases recorded during the study period.

7.5.2 Including the mosquito data in the modelling

The aim of including mosquito surveillance data in the models was to determine if information about mosquito density could increase the sensitivity of the models. Mosquito data were available for only nine of the fourteen SLAs (discussed in Chapter 3, see Figure 7.7 for an illustration of these SLAs). Given this, the results of the
multivariable regression models represent an indicative test of the value of using mosquito data to increase the predictive skill of climate and environment data.

The mosquito data for the months of November and December were derived from graphs of the mean number of *Ae. camptorynchus* mosquitoes trapped per night each month over the study period. In recognition of the imprecision of the data (some of the graphs were presented differently and values had to be estimated), the mosquito density variables were categorised as follows:

- 1 (0 - 99 mosquitoes)
- 2 (100 - 199 mosquitoes)
- 3 (200 - 299 mosquitoes)
- 4 (≥ 300 mosquitoes)

### 7.5.3 Accounting for the influence of tidal height

In terms of how tide height should be managed as a variable in analysis, several methods were examined. Lindsay and others (1989) considered the association between annual mean sea level and outbreak years over a six year period, and reported that mean sea levels were 7 cm below the long-term average in one non-outbreak year, and 5.5 cm above the average in another. They did not report the associations for each single year in the period, however, and the strength and reliability of these observations are not clear. In the same study, they also considered the influence of daily tide heights, and reported that “a number of higher daily tide heights were recorded” in the major outbreak year of 1988/89.

In the 1998 report of Western Australian arboviral surveillance activities (Lindsay *et al.* 1998), tide heights of over 100 cm at the Mandurah coastal tidal gauge (the northern part of the Mandurah SLA) from early October onwards were identified as being a critical factor for the breeding of large numbers of *Ae. camptorynchus* and or *Ae. vigilax*. In the 2001 report of the same activities, tide heights of approximately 110 cm at Fremantle were considered sufficient to inundate at least 60% of the salt-marsh in the Mandurah region, while heights of 87 and 90 cm at Bunbury were considered sufficient to inundate salt-marsh areas of the Bunbury and Harvey SLAs (Broom *et al.* 2001).

In summary, as even one very high tide can be sufficient to inundate a salt-marsh area, the maximum daily tide height in a month (“absolute tide height”) was derived from the
Figure 7.7  Statistical Local Areas within the Southwest study area for which mosquito density data were available (coloured areas), and the approximate distribution of the mosquito trapping areas.
tidal data (obtained from the Coastal Management Branch of the Maritime Division of Transport, Western Australia). The threshold at which tide heights cause widespread salt-marsh inundation appears to vary by location, probably due to different coastal topography, and therefore different absolute tide height thresholds were explored. Finally, the mean tide height in a month was derived ("mean tide height"), and averages of monthly means, and annual mean tide height, were explored as indicators of prolonged coastal inundation conducive to mosquito breeding.

**7.5.4 Reduced data for the late warning model**

In the Murray study area, data for the months of July through to February inclusive (months 1-8) were used to predict the occurrence for epidemics. In the Southwest study area, however, notifications on average commenced and peaked earlier in the RRv year. For the period overall, 79% of notifications in this study area were made up to the end of February (range 68-92% for individual years). By the end of January, however, only 56% of cases had been notified (in 1995/96, when the largest number of epidemics were recorded, only 42% of cases were notified by the end of January). It follows that a prediction of an epidemic at the end of February in the Southwest would be of little use, as nearly 80% of infections (on average) would have already occurred. A late warning model that provided a prediction at the end of January would be in advance of some 40-50% of infections (allowing for a lag in the incubation period). Thus, for this study area the late warning model contained climate, environment, and mosquito variables for the months of July to January (1 to 7) only.

**7.6 Hypotheses**

Based on the published observations to date, the following conclusions were drawn:

- First, the low level of RRv notifications reported in the human population in the three years preceding the large outbreak of 1995/96 suggests the likelihood of a build-up in the number of non-immune vertebrate hosts in the area. Given suitable climatic conditions, a large susceptible host population would have enabled a very large amount of virus to circulate within the natural cycle in 1995/96, resulting in a spill-over of infection into human populations.

- Second, *Ae. camptorhynchus* (and possibly *An. annulipes*) appears to be responsible for initiating and maintaining virus in the vertebrate host population in the late winter to
The Southwest study area spring period. In years when rainfall is above average and tidal conditions are higher than usual, or both, *Ae. camptorynchus* populations may persist into the summer months, providing an opportunity for the infection of summer populations of *Ae. vigilax* (and thus of humans). Alternatively, *Ae. camptorynchus* populations that persist into summer may be responsible on their own for infecting human populations. Either way, climatic conditions in late spring and summer that enable the survival of *Ae. camptorynchus* appear to be a critical predisposing factor for an epidemic. Such conditions are likely to include above average rainfall (in late spring to early summer) and high tides (also in late spring to early summer).

Third, the role of temperature is unclear, but spring and summer temperatures can both be expected to play a role in providing a suitable environment for the survival of adult female mosquitoes. High summer temperatures are linked to increased mosquito death; yet mosquito longevity is enhanced as temperatures increase.

Drawing on these points, four hypotheses were examined:

1. That rainfall in late spring to early summer is positively associated with epidemics.
2. That tide height in late spring to early summer is positively associated with epidemics.
3. That rainfall and tide height or both in the late spring to summer months of the year preceding the epidemic year are negatively associated with epidemics.
4. That early summer (December) *Ae. camptorynchus* density is positively associated with epidemics.

### 7.7 Results

#### 7.7.1 Single variable analysis

**7.7.1.1 Effect of rainfall**

The number of rain days and total rainfall in the late winter (July) were positively associated with epidemics (Figure 7.8), although there was no significant relationship with either of these variables in the following early spring months (August and September). Rain days in November and December were positively associated with
Figure 7.8 Odds ratios for the explanatory variables regressed against epidemics of RRv disease in the Southwest. Odds ratios are on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at P<0.05.

- **Total rainfall**
- **Number of rain days**
- **Southern Oscillation Index**
- **Sea surface temperature**
**Figure 7.9** Odds ratios for the explanatory variables regressed against epidemics of RRv disease in the Southwest. Odds ratios are on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at $P<0.05$.

- **Mean maximum temperature**
- **Absolute maximum temperature**
- **Mean minimum temperature**
- **Absolute minimum temperature**
Figure 7.10  Odds ratios for the explanatory variables regressed against epidemics of RRV disease in the Southwest. Odds ratios are on the Y-axis, and months 1 (July) to 8 (February) on the X-axis. Bolded columns indicate significance at P<0.05.
Figure 7.11 Monthly absolute tide heights, recorded at Bunbury, Western Australia (July 1991 to June 1999).

Data source: Coastal Management Branch, Maritime Division of Transport, Western Australia.
epidemics, and December (early summer) had the strongest and most significant association (OR=1.50, CI=1.23-1.82, P<0.000, R²=18%). There was no significant relationship between either rainfall variable and epidemics for late summer (January and February).

**Derived variables**

There was a statistically significant negative relationship between one year lagged rain days/total rainfall (August through to November) and epidemics. The strongest effect was for August rain days (OR=0.95, CI=0.92-0.97, P<0.000, R²=22%). An average across August to November provided a more reliable biological summary of the effect of the previous year’s spring rainfall on the build-up of large numbers of mosquitoes (rain days: OR=0.70, CI=0.56-0.88, P<0.002, R²=10%).

**7.7.1.2 Effect of the SOI and SST**

October, November, and December SST were negatively associated with epidemics, and the combination of October-November was the best predictor (OR=0.63, CI=0.40-0.99, P<0.045). The SOI for the months of September and October was a significant positive predictor of epidemics (Figure 7.8), although the relationship was weak (August: OR=1.05, CI=1.00-1.10, P<0.04; and September: OR=1.05, CI=1.01-1.10, P<0.016).

**7.7.1.3 Effect of temperature**

With the exception of absolute maximum temperature, there was a general pattern of high temperatures in winter and spring being associated with epidemics - although not all odds ratios were significant (Figure 7.9 and 7.10). There was no significant relationship between late spring/early summer temperatures and the occurrence of epidemics (November through to January).

There was no significant relationship between absolute maximum temperatures in late winter/early spring and the occurrence of epidemics. In November and December, however, low absolute maximums were significantly associated with epidemics. This relationship reversed in January and February (late summer), and high absolute maximums had a strong effect on epidemic occurrence (February: OR=1.79, CI=1.32-2.41, P<0.00, R²=19%). All other temperature variables were also positively associated with epidemics in February (late summer).
7.7.1.4 Effect of tides

There was a significant relationship between high absolute tide height in the months of July to December (with the exception of September) and epidemics (Figure 7.10). The strongest effects were for the months of October through to December (October-November: OR=1.22, CI=1.12-1.33, \( P<0.000, R^2=28\% \)). High mean tides for all months in the study (July to January) also had a significant effect on the occurrence of epidemics. The strongest effect was for December (OR=1.24, CI=1.11-1.37, \( P<0.000, R^2=20\% \)). There were not enough observations to test whether threshold levels in the tide heights were statistically significant.

Derived variables

I derived one year lagged variables for mean and absolute tide height. Figure 7.11 is a time series of the derived monthly tide heights, recorded at Bunbury. The highest absolute tide heights were generally recorded in the July of each year. High absolute tides for the months of July to December (except September) were significantly associated with epidemics. The strongest effects were with averages for the months of July-December (OR=1.11, CI=1.05-1.16, \( P<0.000, R^2=14\% \)), and October-December (OR=1.18, CI=1.09-1.27, \( P<0.000, R^2=21\% \)). Average annual tide height was not significantly associated with epidemics.

7.7.1.5 Effect of mosquito trap data

November and December mosquito density variables were very strongly associated with the occurrence of epidemics (Table 7.5). November mosquito density for the previous year was negatively associated with epidemics (at the higher \( P<0.2 \) threshold for single variable analysis).
Table 7.5  Single variable logistic regression model results for mosquito density in the Southwest study area: odds ratios (OR), standard errors (SE), and confidence intervals (CI).*

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>SE</th>
<th>P value</th>
<th>95% CI</th>
<th>R²***</th>
</tr>
</thead>
<tbody>
<tr>
<td>November mean mosquito density</td>
<td>5.96</td>
<td>3.60</td>
<td>0.003</td>
<td>1.82-19.81</td>
<td>29%</td>
</tr>
<tr>
<td>December mean mosquito density</td>
<td>13.21</td>
<td>8.05</td>
<td>0.000</td>
<td>4.0-43.29</td>
<td>57%</td>
</tr>
<tr>
<td>November x December mean mosquito density</td>
<td>1.97</td>
<td>0.32</td>
<td>0.000</td>
<td>1.42 - 2.73</td>
<td>58%</td>
</tr>
<tr>
<td>Lagged November mean mosquito density</td>
<td>0.633</td>
<td>0.22</td>
<td>0.200</td>
<td>0.31-1.27</td>
<td>3%</td>
</tr>
</tbody>
</table>

* Model results for the mosquito data are based on a reduced data set (only nine SLAs out of a total of fourteen).

** R² = Percentage of variance explained by the model.

7.7.2 Early warning models (months 1-5)

7.7.2.1 Southwest early warning model

Epidemics in the Southwest area were best predicted by (i) an average of rain days from August to December in the preceding year, (ii) late spring sea surface temperature, and (iii) late spring rainfall and tide variables. The logistic regression model with the best fit had the form:

\[
\log \left[ \frac{P}{1-P} \right] = a + \beta_1 (\text{lag RainDays}_{2-6}) + \beta_2 (\text{SST}_{45}) + \beta_3 (\text{Tide}_{45}) + \beta_4 (\text{RainDays}_{5})
\]

Table 7.6 shows the predictor variables, their odds ratios, 95% confidence intervals, coefficient estimates, and standard errors.

Table 7.6  Southwest early warning logistic regression model: odds ratios (OR), 95% confidence intervals (CI), coefficients, and standard errors (SE).

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>P value</th>
<th>Coeffs</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain days (Aug-Dec) - one year lag</td>
<td>0.80</td>
<td>0.59 - 1.07</td>
<td>0.134</td>
<td>-0.221</td>
<td>0.12</td>
</tr>
<tr>
<td>sea surface temperature (Oct-Nov)</td>
<td>9.94</td>
<td>1.96 - 50.34</td>
<td>0.006</td>
<td>2.296</td>
<td>8.22</td>
</tr>
<tr>
<td>absolute tide height (Oct-Nov)</td>
<td>1.75</td>
<td>1.24 - 2.46</td>
<td>0.001</td>
<td>0.559</td>
<td>0.31</td>
</tr>
<tr>
<td>rain days (Nov)</td>
<td>1.34</td>
<td>1.05 - 1.69</td>
<td>0.017</td>
<td>0.289</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Model R² (percentage of variance explained by model) = 48%
Hosmer and Lemeshow goodness-of-fit test: 0.72 (prob > chi² = 0.99)

Rain days from August to December of the year preceding the epidemic year were negatively associated with epidemics. Rain days across this six-month period was a better predictor than rain days just for the three months of October to December. Rain days was a better predictor than total rainfall.
Sea surface temperatures for late spring (Oct-Nov) were positively associated with epidemics. Linked to this, late spring precipitation and high tides, which both provide unseasonal water to the area, were significant positive predictors of epidemics. The number of rain days in November was substantially more significant than total rainfall in the same month. The average number of November rain days for the period was 7 (range 2 - 14). Some 36% of epidemics occurred in years with fewer than 7 days of rain in November, compared to 66% of non-epidemics. The best predictor of tidal influence was absolute tide height (Oct-Nov average). The period mean of October-November absolute tide height was 110cm (range 96-124cm). Some 86% of epidemics occurred in years when the October-November absolute tide height was above 110cm, compared to 41% of non-epidemics. Some 54% of epidemics occurred in the year of the highest recorded October-November absolute tide height (124cm in 1995/96).

Cross-validation

Table 7.7 shows the validation test results for individual years. The model’s sensitivity (percentage of epidemics correctly predicted) was 64%, and the specificity (percentage of non-epidemics correctly predicted) was 96%. The accuracy (number of epidemics and non-epidemics correctly predicted) was 89%.

There was almost no change to the predictive skill of the model when the second outcome variable was tested (Epidemic Definition 2). The sensitivity remained 64% and the specificity was 92%.
Table 7.7 Sensitivity, specificity, positive predictive values, and negative predictive values for the early warning model in the Southwest.

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of epidemics</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-epidemics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91/92</td>
<td>2</td>
<td>12</td>
<td>50%</td>
<td>92%</td>
<td>50%</td>
</tr>
<tr>
<td>92/93</td>
<td>0</td>
<td>14</td>
<td>-</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>93/94</td>
<td>0</td>
<td>14</td>
<td>-</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>94/95</td>
<td>0</td>
<td>14</td>
<td>-</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>95/96</td>
<td>12</td>
<td>2</td>
<td>100%</td>
<td>0%</td>
<td>86%</td>
</tr>
<tr>
<td>96/97</td>
<td>4</td>
<td>10</td>
<td>25%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>97/98</td>
<td>3</td>
<td>11</td>
<td>0%</td>
<td>91%</td>
<td>0%</td>
</tr>
<tr>
<td>98/99</td>
<td>1</td>
<td>13</td>
<td>0%</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>All years</td>
<td>22</td>
<td>90</td>
<td>64%</td>
<td>96%</td>
<td>78%</td>
</tr>
</tbody>
</table>

1 Number of epidemics | non-epidemics that occurred each year in the Southwest.

7.7.2.2 Southwest sub-region early warning model, with mosquito data

Mosquito data were only available for nine of the fourteen SLAs in the Southwest study area (detailed in Chapter 3). To determine if sensitivity could be improved by mosquito density data, a model was developed for the nine SLAs (called the “Southwest sub-region” model), that used the same predictor variables as the Southwest early warning model. Of the candidate mosquito density variables for the period, lagged November mosquito density was the most plausible and the best predictor. The variables sea surface temperature (Oct-Nov), absolute tide height (Oct-Nov), and rain days (Nov) remained the same as in the Southwest model. One year lag rain days (Aug-Nov), the other variable from the Southwest model, had a moderately high correlation with lagged November mosquito density (0.43). In addition, it added nothing to the model (P>0.266) and was dropped. The logistic regression model with the best fit had the form:

$$\log\left[\frac{P}{1-P}\right] = \alpha + \beta_1(\text{lagMosq}_5) + \beta_2(\text{SST}_45) + \beta_3(\text{Tide}_45) + \beta_4(\text{RainDays}_5)$$

Table 7.8 shows the results of the early warning model for the Southwest sub-region with mosquito data. One year lag November mosquito density was negatively associated with epidemics, although this was not statistically significant (P<0.259). Even so, the variable was kept in the model to assess its contribution.
Table 7.8  Southwest sub-region* early warning logistic regression model, with mosquito density data: odds ratios (OR), 95% confidence intervals (CI), coefficients, and standard errors (SE).

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>P value</th>
<th>Coefs</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>mosquito density (Nov) - one year lag</td>
<td>0.56</td>
<td>0.19 - 1.61</td>
<td>0.287</td>
<td>-0.572</td>
<td>0.30</td>
</tr>
<tr>
<td>sea surface temperature (Oct-Nov)</td>
<td>7.83</td>
<td>0.61 - 99.0</td>
<td>0.113</td>
<td>2.058</td>
<td>10.17</td>
</tr>
<tr>
<td>absolute tide height (Oct-Nov)</td>
<td>2.02</td>
<td>1.16 - 3.52</td>
<td>0.013</td>
<td>0.704</td>
<td>0.57</td>
</tr>
<tr>
<td>rain days (Nov)</td>
<td>1.60</td>
<td>1.03 - 2.49</td>
<td>0.036</td>
<td>0.472</td>
<td>0.36</td>
</tr>
</tbody>
</table>

* The Southwest sub-region comprises 9 of the 14 Southwest SLAs.
Model R^2 (percentage of variance explained by model) = 64%
Hosmer and Lemeshow goodness-of-fit test: 0.77 (prob > chi2 = 0.93)

Cross-validation

The validation test\textsuperscript{15} results (Table 7.9) for the validated Southwest sub-region model with mosquito data was not simple to estimate. No lagged mosquito data were available for 1997/98 (because no mosquito data were available for 1996/97). In addition, there was no estimate of the probability of epidemics for 1996/97 because the mosquito data predicted non-epidemics completely. Thus, predictions could not be formally estimated for two of the years in the study period. The validation test is based on estimations for ten of the sixteen epidemics that occurred. The accuracy (were the additional six epidemics correctly or incorrectly predicted) is also presented. With the available data, the model had a sensitivity of 90% (possible range 56 - 94%). The specificity was 88% (possible range 69 - 91%). Overall accuracy was 88% (possible range 66 - 91%). A slight decrease in sensitivity was noted with the second outcome variable (90% to 80%).

\textsuperscript{15} Validation Test. \textit{Sensitivity}: the proportion of epidemics that the model correctly identifies as epidemics. \textit{Specificity}: the proportion of non-epidemics that the model correctly identifies as non-epidemics. \textit{Positive predictive value}: the likelihood that an epidemic predicted by the model really is an epidemic. \textit{Negative predictive value}: the likelihood that a non-epidemic predicted really is a non-epidemic. \textit{Accuracy}: the percentage of epidemics and non-epidemics correctly predicted. Except where otherwise noted, a cut-off probability of 0.5 was used to define an epidemic.
Table 7.9  Sensitivity, specificity, positive predictive values, and negative predictive values for the early
warning model in the Southwest sub-region – with mosquito data.

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of epidemics</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>1</td>
<td>8</td>
<td>100%</td>
<td>33%</td>
<td>20%</td>
</tr>
<tr>
<td>92/93</td>
<td>0</td>
<td>9</td>
<td>–</td>
<td>100%</td>
<td>–</td>
</tr>
<tr>
<td>93/94</td>
<td>0</td>
<td>9</td>
<td>–</td>
<td>100%</td>
<td>–</td>
</tr>
<tr>
<td>94/95</td>
<td>0</td>
<td>9</td>
<td>–</td>
<td>100%</td>
<td>–</td>
</tr>
<tr>
<td>95/96</td>
<td>8</td>
<td>1</td>
<td>100%</td>
<td>0%</td>
<td>89%</td>
</tr>
<tr>
<td>96/97</td>
<td>3</td>
<td>6</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>97/98</td>
<td>3</td>
<td>6</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>98/99</td>
<td>1</td>
<td>8</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>All years</td>
<td>16</td>
<td>56</td>
<td>(90% )</td>
<td>(88%)</td>
<td>(64%)</td>
</tr>
</tbody>
</table>

1 Number of epidemics | non-epidemics each year in the Southwest sub-region.
2 Mosquito data were not available for this year, and Stata did not provide probability estimates.
3 The mosquito variable predicted non-epidemics completely, and Stata did not provide probability
   estimates.
4 Model results are presented and (the range of possible results had the data for Years 6 and 7 been
   available).

7.7.3 Late warning models (months 1-7)

7.7.3.1 Southwest late warning model

I added candidate variables for December and January to determine if additional climatic and
environment data for these months could improve the prediction of epidemics. None of the variables improved the prediction of epidemics above the level already achieved by the early warning model (i.e., 64% sensitivity).

7.7.3.2 Southwest sub-region late warning model, with mosquito data

Mosquito density variables for December and January were added to the Southwest sub-region early warning model. Of the available mosquito density variables, November-December mosquito density was significant, and the best predictor. Absolute tide height (Oct-Nov), sea surface temperature (Oct-Nov), and rain days (Nov) were non-significant and were dropped from the model. Although highly non-significant, the other variable in the early warning model (one year lag rain days for August to December) was retained.
because of its essential role in the transmission cycle. The logistic regression model with the best fit had the form:

$$\log\left(\frac{P}{1-P}\right) = \alpha + \beta_1(\text{lagRainDays}_{2-6}) + \beta_2(\text{Mosq}_{56})$$

Table 7.10 shows the results of the Southwest sub-region late warning model with mosquito data.

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
<th>P value</th>
<th>Coeffs</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain days (Aug-Dec) - one year lag</td>
<td>1.19</td>
<td>0.63 - 2.27</td>
<td>0.592</td>
<td>0.176</td>
<td>0.39</td>
</tr>
<tr>
<td>mosquito density (Nov-Dec)</td>
<td>2.16</td>
<td>1.32 - 3.51</td>
<td>0.002</td>
<td>0.767</td>
<td>0.24</td>
</tr>
</tbody>
</table>

* The Southwest sub-region comprises 9 of the 14 Southwest SLAs.

Model $R^2$ (percentage of variance explained by model) = 58%

Hosmer and Lemeshow goodness-of-fit test: 8.70 (prob > chi2 = 0.37)

One year lag November mosquito density could have been a better predictor than one year lag rain days (Aug-Dec). However, the variable predicted non-epidemics completely and I dropped it from the model.

**Cross-validation**

The limitations discussed for the Southwest sub-region early warning model (with mosquito data) regarding the estimation of the model's sensitivity apply to this model as well. The available data show a 5% reduction in sensitivity (to 85%) from the early warning model. However, the sensitivity for this late warning model was estimated on the basis of seven years data, whereas the early warning model had only six years. Consequently, these results are likely to be more precise. The specificity and accuracy were higher than for the early warning model (98% and 95% respectively). Table 7.11 shows the results of the validation test.
Chapter 7 – The Southwest study area

Table 7.11  Sensitivity, specificity, positive predictive value, and negative predictive value for the early warning model in the Southwest sub-region – with mosquito data.

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of epidemics</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/92</td>
<td>1</td>
<td>8</td>
<td>0%</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>92/93</td>
<td>0</td>
<td>9</td>
<td>-</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>93/94</td>
<td>0</td>
<td>9</td>
<td>-</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>94/95</td>
<td>0</td>
<td>9</td>
<td>-</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>95/96</td>
<td>8</td>
<td>1</td>
<td>100%</td>
<td>0%</td>
<td>89%</td>
</tr>
<tr>
<td>96/97</td>
<td>3</td>
<td>6</td>
<td>67%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>97/98</td>
<td>3</td>
<td>6</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>98/99</td>
<td>1</td>
<td>8</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

All years 16 | 56 85% ³ 98% 92% 96% (69 - 88%) (87 - 98%) (61 - 93%) (90 - 96%)

1 Number of epidemics | non-epidemics each year in the Southwest sub-region.
2 Mosquito data were not available for this year, and Stata did not provide probability estimates.
3 Model results are presented and (the range of possible results had the data for Year 6 been available).

7.7.4 Summary

In relation to the four hypotheses outlined in Section 7.6, the following were found:
1 The number of rain days in late spring was a positive predictor of epidemics.
2 Absolute tide heights in late spring were positively associated with epidemics.
3 Rainfall (specifically the number of rain days) in the spring to early summer months of the pre-epidemic year was negatively associated with epidemics. Tide heights in the same period did not predict epidemics.
4 Early summer *Ae. camptorynchus* density was positively associated with epidemics in the Southwest sub-region. However, the combination of November and December density was the best predictor. November mosquito density in the pre-epidemic year was negatively associated with epidemics.

Table 7.12 provides a summary of the variables in each of the models, and Table 7.13 gives the sensitivity analyses for the models (all years). As in the Murray study area, all the models predicted non-epidemics better than epidemics. Epidemic prediction for the Southwest (all 14 SLAs) did not increase with the addition of climate and environment data for the months of December and January. In the Southwest sub-region, the addition of mosquito density data dramatically improved the sensitivity of the early warning model, without prejudicing the specificity. Although the sensitivity appeared to drop for
the late warning model, the confidence interval estimates were closer; the specificity and overall accuracy was also higher than for the early warning model.

Table 7.12 Summary table: predictor variables for the early warning and late warning models for the Southwest study area.

<table>
<thead>
<tr>
<th>Early Warning Models</th>
<th>Late Warning Models</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Southwest</strong>¹</td>
<td>• no improvement on the early warning model</td>
</tr>
<tr>
<td></td>
<td>• rain days (Aug-Dec) - one yr lag</td>
</tr>
<tr>
<td></td>
<td>• sea surface temperature (Oct-Nov)</td>
</tr>
<tr>
<td></td>
<td>• absolute tide height (Oct-Nov)</td>
</tr>
<tr>
<td></td>
<td>• rain days (Nov)</td>
</tr>
<tr>
<td></td>
<td>• mosquito density (Nov - Dec)</td>
</tr>
<tr>
<td><strong>Southwest sub-region</strong></td>
<td>• sea surface temperature (Oct-Nov)</td>
</tr>
<tr>
<td></td>
<td>• absolute tide height (Oct-Nov)</td>
</tr>
<tr>
<td><strong>Southwest sub-region + mosquito</strong>²</td>
<td>• rain days (Aug-Dec) - one yr lag</td>
</tr>
<tr>
<td></td>
<td>• mosquito density (Nov-Dec)</td>
</tr>
</tbody>
</table>

¹ Southwest study area: 14 SLAs.
² Southwest sub-region: 9 SLAs.

Table 7.13 Summary table: sensitivity, specificity, positive predictive values, negative predictive values, and accuracy of the early warning and late warning models for the Southwest study area, and the Southwest sub-region with mosquito data.

<table>
<thead>
<tr>
<th>Predictive Skill</th>
<th>Early Warning Models</th>
<th>Late Warning Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Southwest</strong>¹</td>
<td><strong>Southwest sub-region + mosquito</strong>²</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>64% (56 - 94%)</td>
<td>90% (69 - 88%)</td>
</tr>
<tr>
<td>Specificity</td>
<td>96% (69 - 91%)</td>
<td>88% (87 - 98%)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>78% (35 - 75%)</td>
<td>64% (61 - 93%)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>91% (84 - 98%)</td>
<td>97% (90 - 96%)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>89% (66 - 91%)</td>
<td>88% (83 - 96%)</td>
</tr>
</tbody>
</table>

¹ Southwest study area: 14 SLAs, 22 epidemics and 90 non-epidemics.
² Southwest sub-region + mosquito data: 9 SLAs, 16 epidemics and 56 non-epidemics. Data unavailable for 2 of the 8 years. Model results are presented and (the range of possible results had all the data been available).
³ Southwest sub-region + mosquito data: 9 SLAs, 16 epidemics and 56 non-epidemics. Data unavailable for 1 of the 8 years. Model results are presented and (the range of possible results had all the data been available).
7.8 DISCUSSION

These findings suggest that the methods developed and used to predict epidemics of RRv disease in the Murray study area could be applied to other RRv epidemic regions of Australia with some success. For the Southwest study area as a whole, climate and environment data on their own were able to predict epidemics with moderate skill (64%) from early in December. A low number of rain days in the spring/early summer of the pre-epidemic year, combined with suitable climatic conditions by November (high sea surface temperatures and high tide heights for October-November, and a high amount of rainfall for November) best predicted epidemics. There was a consistent relationship between these findings and previous research.

Climate and environment data for December and January made no increase in predictive skill, over and above that which had already been provided by the variables for July to November. This suggests that the climatic and environmental preconditions for an epidemic were already established by the end of November, and that - given the available data - conditions in December and January were less important in determining an event in this study area.

In addition to the predictive skill provided by climate and environment data, mosquito surveillance data contributed significantly as an RRv disease early warning indicator. Mosquito density data for the November of the pre-epidemic year increased the sensitivity of the early warning model by an estimated 26% (range -8% to 30%). For the late warning model, November-December mosquito density increased sensitivity by 21% (range 5% to 24%). These findings support the value of mosquito surveillance data for use as an early warning indicator (discussed further in Chapter 8).

7.8.1 Limitations of the study

The following factors need to be considered in an evaluation of the results.

- Several limitations that were noted for the Murray study area are also relevant to the findings from the Southwest. These are: the possibility that some RRv infections were spatially mis-classified; the short time series of the disease data; and the limits of the cross-validation method.

- The tidal data were available for a fixed point only (Bunbury) and values were applied to all SLAs. It is likely that tide heights vary along the coast to the north and
south of this point, although there is no information available to guide the direction or magnitude of these possible differences. As noted earlier, the contribution of tidal inundation to the initiation of RRv disease epidemics in the southern SLAs is uncertain, and warrants further research. More specific tailoring of tidal information to relevant SLAs would improve the accuracy of the results.

- Mosquito data proved to be a very valuable indicator of RRv epidemics in this study area, in addition to climate data. However, several data quality issues suggest that these results should only be considered an approximate indication of the additional sensitivity that mosquito data can contribute to prediction. Firstly, mosquito data were lacking for one year of the study, which meant that epidemics for that year were not able to be predicted. Secondly, mosquito data were not collected in all of the SLAs. For the purpose of this analysis, I assumed that mosquito density in adjacent SLAs would be proportionately similar. But as trap levels within a single SLA can vary (e.g. Lindsay et al. 1998), further research is needed to establish whether this assumption is reasonable. Thirdly, the method used to determine the mosquito density for each SLA was simplistic, and some inaccuracy may have been introduced in the process. Furthermore, the mosquito data were obtained from three sources, and I cannot be certain that the method used to determine mosquito density was the same for each source. I used a ranking method, rather than individual density values, in an attempt to minimise possible measurement differences.

7.8.2 The influence of rainfall

As was the case in the Murray study area, low rainfall in the pre-epidemic year was a significant predictor in the models, presumably for the same reasons - that is, the recruitment of susceptible vertebrate hosts. A low number of rain days across the five months from late winter to early summer would inhibit the breeding of large numbers of mosquitoes, and thus reduce the transmission of virus within the natural cycle. As a consequence, there would be more susceptible vertebrate capable of sustaining the transmission cycle in the following year.

Mackenzie and others (2000) have noted that “salt marsh mosquito breeding can be exacerbated by rainfall in the absence of, or in addition to, tidal inundation”. The combination of rain days and absolute tide height in November (the end of spring) were predictors in all models. However, November rain days made only a small contribution to the epidemic prediction of the models, which suggests that rainfall did not add as
much as tidal inundation to the creation of ideal breeding conditions capable of sustaining epidemic numbers of mosquitoes.

7.8.3 The influence of tidal inundation and sea surface temperature

These results showed that tidal height (and, by inference, tidal inundation of salt marsh mosquito breeding areas) was positively associated with epidemics, and the variable October-November tidal height was a predictor in all the models. Previous research has observed tidal height to be associated with RRv epidemics in the same area (Lindsay et al. 1996c, Lindsay et al. 1998). Of the tidal variables, monthly absolute height was consistently a better predictor than monthly average height.

Anomalous cooling of the sea water in the eastern Pacific (a La Niña) typically corresponds to a warming of the sea waters, and an increase in the mean sea level and high tides, around Western Australia (Allan et al. 1996). The El Niño phase presents the reverse sea level pattern. From 1991/92 to 1994/5 there were five consecutive years of El Niño events (Climate Prediction Center 2001). Absolute tide recordings for October-November in those years ranged from 96-110 cm. During October-November 1995 there was a "weak strength" La Niña (Climate Prediction Center 2001), and the highest tide height for the period was recorded (124 cm). In that year, twelve of the fourteen SLAs in the Southwest sub-region recorded epidemics. In the following year (1996/97), the second highest absolute tide height for the same two months was recorded (120 cm). In that year, only four epidemics were recorded. Of these, two were in SLAs that had not recorded epidemics in the previous year.

This observation suggests two things. First, the link between sea surface temperature and tide height in this region is strong. Second, high tides of their own are not sufficient to generate epidemics of RRv disease. Consecutive years of such events do not appear to allow time for the accumulation of large numbers of susceptible vertebrate hosts.

In the single variable analysis, sea surface temperature for October-November ("SST45") demonstrated a significant negative association with epidemics. As noted, this is presumably the result of the teleconnected influence of La Niña, where local sea warming occurs and rainfall and tide heights increase. However, when SST45 was entered into the early warning model with both rain days (for Nov) and absolute tide height (for Oct-Nov), the association became positive (SST45 OR= 9.10, P<0.006). One conceivable
explanation is that, over and above the influence of SST45 on tide height and rainfall (which was negative), local cooling of the ocean has a beneficial impact on mosquito breeding. This may be explained by the influence of El Niño conditions on temperature. Clouds slow the rate at which the earth warms during the day and cools at night. When sea surface temperatures in the eastern Pacific are negative (causing local warming and increased rainfall), average monthly temperatures in Australia are typically reduced (Partridge 1994). The converse is also true - local cooling of the oceans increases the likelihood of warmer temperatures. All temperature variables for October (month 4) were positively associated with epidemics, and with the SST45.

Warmer temperatures are likely to increase the probability of RRv disease epidemics by lengthening the life span of the mosquitoes and increasing the extrinsic incubation period of the virus. Gagnon and others (2001) noted a similar phenomenon in relation to dengue epidemics in northern South America (epidemics occurred more frequently with warmer temperatures and drier conditions). They queried whether the magnitude of temperature anomalies during an El Niño would be large enough to cause dengue epidemics without the contribution of rainfall variations (although they further noted that Patz (1998) concluded dengue epidemics could increase worldwide with a temperature increase of only 1°C). In the present study, warm temperatures were accompanied by both rainfall and high tides (which were positive and significant predictors in the model).

The sea surface temperature variable provides additional value as an indicator in an early warning system. Changes in the El Niño “state” of the eastern Pacific Ocean can be predicted, on average, from winter onwards in Australia (Stone et al. 1996). This, coupled with knowledge of the preceding year’s rainfall and mosquito density, provides several earlier indicators of the probability of an impending epidemic season.

7.8.4 Contribution of the mosquito data to model sensitivity

In the early warning model, climate and environment data on their own were able to predict 64% of epidemics, and mosquito density information appeared to add additional sensitivity. The results of the late warning model suggests, with more confidence, that epidemics in the Southwest can be confirmed with high sensitivity and specificity by early February using mosquito trapping data alone.
While rainfall and tidal inundation are probably the ultimate factors determining the occurrence of epidemics in this study area, mosquito density is the proximate factor. Thus, rainfall and tide variables operated as surrogate measures of mosquito density in the models. As expected, there was a moderately high and positive correlation between mosquito density and both these variables (e.g. tide height (0.58) and rain days (0.63)). High mosquito numbers in the previous year would imply active virus circulation and hence maintenance of high immunity levels in vertebrate host populations. This is consistent with a negative association between lagged mosquito densities and epidemics, as found.

Although November mosquito density for the previous year appeared to provide additional sensitivity to the early warning model, in an ideal world it could probably be improved upon. Mosquito density data for the whole spring/early summer period (if these were available) may be better at describing the level of mosquitoes – and hence of viral activity in the pre-epidemic year – than lagged November density on its own. This speculation is supported by the relationship between the rain day variables from the same period: lagged November rain days had a significant negative effect on epidemic occurrence (OR=0.79, P<0.015). The negative effect increased when information about all the spring months were added (OR=0.70, P<0.002).

7.8.5 Comparison with the Murray study areas

There were some consistent patterns in the predictor variables between the two study areas - which is not surprising, given that two of the three RRv bioclimatic regions are classified as temperate climatic zones. The importance of low levels of rain days in the pre-epidemic year suggests that an extended period (i.e., more than one year) is needed to generate large enough numbers of susceptible hosts to support an epidemic. Unfortunately the time series was too short to provide evidence of whether a longer period is required. Observations from arid Australian environments suggest that in such areas – where vertebrate host breeding is profoundly linked to rainfall, and where rainfall is low and variable - the RRv cycle is three to four years at minimum.

Temperatures did not emerge as useful predictor variables in the Southwest, although they were in the Murray. This is not to say that temperature is not an important factor driving epidemics: in single variable analysis all temperature variables (except absolute maximum temperatures) showed a positive association with epidemics. Rather, other
Chapter 7 - The Southwest study area

Predictor variables (possibly SST45) appeared to contribute more to the models than temperature.

Why it was not possible to develop a stronger prediction model using climate data alone for the Southwest study area is not clear, but findings from the analysis suggest a couple of reasons. The method used to define the boundaries of the Southwest study area did not account for a particular regional factor that fundamentally determines mosquito ecology in the northern part of the study area - tidal inundation of salt marshes. In the Murray study area the principal drivers of mosquito breeding and survival are rainfall and temperature. These drivers were two of the three criteria used to define homogeneous RRv bioclimatic regions. In the Southwest, no broad-scale information on wetland ecology was used (or available). If it had been, the shape of the region may have been different. The northern nine SLAs of the Southwest sub-region are in or closely adjacent to the region referred to as the Swan Coastal plain, where mosquito breeding is strongly affected by tidal inundation. There is no evidence that tide height influences the mosquito breeding cycle in the Southern SLAs, where wetlands diminish and are replaced by escarpment and coastal heath.

In support of this, there appears to be a difference in the seasonal timing of epidemics between the northern and southern SLAs in the Southwest study area. Case follow-up data for 1987 to 2002 show that the largest percentage of cases in the Southwest sub-region (the north) occurred in January (Michael Lindsay pers. comm. 2002), and some 72% of cases had occurred by the end of January. However, in the southern SLAs cases peaked, on average, a month later in February. This suggests that epidemics are initiated in the north and travel south, when weather conditions in those parts become suitable.

7.8.6 Implications for climate change

Slightly less warming is projected for the south-west of Australia than for the rest of the country. By 2070, summer and autumn temperatures are projected to increase by 1-5°C, and winter and spring temperatures to increase by 1-4°C (CSIRO 2001). Changes in extremely hot temperatures (i.e., days above 35°C), assuming no change in variability, are predicted to increase. Most climate models consistently project increased drying for south-western Australia during the autumn through to spring seasons (-60% to +10% by 2070). This would mean more extreme dry spells. At the same time, global warming is
expected to lead to an increased frequency of extreme precipitation events (even when an overall decrease in rainfall is projected).

The projected overall drying for the study area and the increase in extreme rainfall events could result in a decrease in the frequency of RRv disease epidemics, and an increase in their severity (i.e., more cases in an epidemic year). A longer spacing between epidemics would allow for greater accumulation in the proportion of susceptible vertebrates in the environment. If there was an increase in extreme rainfalls in late spring, this would enhance conditions for the breeding of *Ae. camptorynchus* - other factors being equal.

The average increase in temperatures is likely to be generally favourable for mosquito development in this study area, as elsewhere. An increase in extreme temperature events would support increased viral amplification (and hence an increase in infectious mosquitoes). However, mosquito survival would be dependent on extreme temperatures being accompanied by sufficiently humid conditions.

Given the importance of tide height on the initiation of RRv disease epidemics in this study area, the projections for increases in mean sea level due to climate change are relevant. Projections are for a rise of 9 to 88 cm by 2100 (or 0.8 to 8.0 cm per decade: IPCC 2001b). These suggest that tidal inundations could occur more frequently than at present, which would be a positive factor for mosquito breeding in what are otherwise expected to be dryer summers.
Chapter 8

Epidemic early warning systems

8.1 INTRODUCTION

An early warning system brings together research and policy into a response plan, which aims to minimise the risks to population health of a threat (such as an outbreak of mosquito-borne disease, a natural disaster, or a famine). This chapter presents the core requirements of early warning systems. It considers the feasibility of establishing early warning systems for RRv disease, and uses the models developed for the case study areas (the Murray and the Southwest) to describe the practical issues involved.

The importance of adequate and timely response to the success of an early warning system is often discussed only as an addendum to predictive modelling (Myers et al. 2000). Yet the purpose of an early warning system is to guide decision-making regarding disease preventive actions. The common division of responsibility between those who generate disease prediction techniques, and those who organise a response based on warnings (the "policy-makers"), means that the two primary components - prediction and response - are usually developed in isolation. This split is a potential limitation of many early warning systems. This chapter makes the case for focusing, in an integrated fashion, on both the prediction and response components of early warning systems. It argues for a re-orienting of response towards the individual and population factors that reduce infection, rather than continuing our reliance on the more traditional methods of vector control.
8.2 HISTORICAL DEVELOPMENT OF EARLY WARNING SYSTEMS

Our ability to use our knowledge of Earth's natural systems to predict future risks to health has taken a quantum leap in the past thirty years. In this period, weather forecasting has gradually increased from same-day forecasting to the three day advance forecast, and our understanding of the mechanics and teleconnections of ENSO now provides the capacity for seasonal and annual rainfall forecasting - assumed as recently as the 1970s to be the stuff of fancy not science (Nicholls 2002). The SOI is now a standard tool used by farmers worldwide for drought monitoring and crop yields. The need for famine early warning systems was first recognised in the early 1970s, following droughts in Africa (Glantz 1997), and initial systems were established at the global level in 1974 by the Food and Agricultural Organization (FAO) and USAID (Moseley and Logan 2001). But it was not until after widespread famines in Africa during the mid-1980s that the importance of climate and other information in determining the timeliness of response was recognised, which prompted a boom in the famine early warning system industry (Buchanan-Smith 1997). The majority of these systems operate at the national level, although there are also global, regional and local systems designed to monitor famine and food security.

Famine early warning systems are themselves a subset of disaster early warning systems, which have been developed to warn of other natural disasters such as floods, volcanoes, and cyclones (timely earthquake warnings still remain elusive). Many countries (e.g. Australia, Italy, the United States, China) are now considering the establishment of heat-watch warning systems to give advance warning of heatwaves. Some of the best-developed early warning systems are for plague locust control, which use remote sensing, GIS and decision support tools to identify regions where pests will build up (WHO 2001).

In the area of infectious diseases, work in the United States has focused on the use of dead bird surveillance as an early warning of West Nile virus activity (Eidson et al. 2001). An early warning system for Rift Valley fever based on rainfall has been proposed for Kenya (see for example Linthicum et al. 1999), and the FAO is investigating the value of developing a system in the east African region (Garner and Kalunda 1999). Historically, a highly successful mathematically-based epidemic forecasting system for malaria functioned in the Punjab, India, from the early 1920s until 1947 (Swaroop 1949). Various workers (Gill 1923, Gill 1927, Yacob and Swaroop 1944, Yacob and Swaroop 1945, Yacob
and Swaroop 1946, Swaroop 1949) refined a system that combined climatic factors (such as rainfall) with health and economic factors (the regional spleen index, a measure of ecological vulnerability, and food prices).

Today, the greatest effort in mosquito-borne disease early warning systems is being directed towards the global fight to combat malaria (principally in African nations), organised through the WHO Global Roll Back Malaria programme (WHO 1998b). The WHO has sponsored substantial research into this area, and a framework has been clarified to address epidemic forecasting, early warning, detection, and control for malaria (HIMAL and MARA 1999, WHO 2001). Medium-range climate forecasts (such as rainfall predictions based on El Nino) and real time weather data (rainfall) are being considered as indicators.

8.3 REQUIREMENTS FOR EFFECTIVE EARLY WARNING SYSTEMS

There is wide variation in the meanings of the term "early warning system". Most early warning systems are heavily focused on developing or describing the prediction component of the system (Cox et al. 1999, Myers et al. 2000, Hay et al. 2001). For example, Myers and others (2000) identified the principal components of infectious disease early warning systems to be (i) the routine surveillance of disease, and (ii) the modelling of current and future risk.

Although they provide the basis for an early warning system, disease predictive models do not on their own constitute one. Even when predictive models are comprehensive, and are able to account for physical or biological feedback, climate change, and human adaptation (Committee on Climate 2001), they are still of limited value until they are actually used to direct disease control efforts. A "system" (in the context of mosquito-borne diseases) suggests an interrelation of predictive models, data, procedures, interventions and communication between groups that combine to accomplish a specific task - i.e., the reduction in cases of disease in a region. An early warning "system", to be effective, has to elicit an appropriate societal response (Walker 1989, Davies et al. 1991, Buchanan-Smith 1997, Saidy 1997, Tapscott 1997).

The purpose of an early warning system is to reduce vulnerability and increase preparedness (Committee on Climate 2001). To do this effectively, early warning systems
Chapter 8 – Early warning systems

should integrate the two principal components of disease prediction and response, with the situation-specific constraints for response designed into the development of the prediction model. Requirements for the prediction end are:

- knowledge of disease transmission
- reliable and up-to-date information on the exposure and health outcome
- a model that is accurate, specific, and timely.

Requirements for the response end are:

- a response plan, detailing thresholds for action
- available and effective interventions
- economic assessment of the benefit and affordability of the system
- involvement of all relevant stakeholders in the process
- a communication strategy.

The following sections argue the case for these general requirements, and discuss the possibilities for the application of early warning systems for RRv disease in Australia (focusing on the case study areas).

8.4 Disease prediction

8.4.1 Knowledge of the disease transmission process

A knowledge of the disease cycle (including the influence of climate and environmental factors in modifying this cycle) and the availability of quality exposure and health data are essential pre-conditions for an early warning system (Garner and Kalunda 1999). As a consequence, “operational early warning systems are not yet generally feasible, due to our limited understanding of most relationships between climate and disease” (Committee on Climate 2001)

Not all regions currently affected by RRv disease would be suitable for the development of an early warning system. RRv has a complex transmission cycle that occupies several different ecological zones (Russell 1994). To develop a prediction model, data are needed on the timing of past outbreaks, as well as regional information on the vector and host species involved in transmission, and the environmental and climatic factors that influence their breeding and survival. In many parts of Australia this information is entirely lacking, or not sufficiently detailed. Mosquito trapping and virus isolation, for example, only occur in few parts of the country, and to date no surveillance has been
conducted of RRv host populations. To maximise the use of resources, trap positions are
often selected with the aim of obtaining information on multiple disease vectors (usually
with a priority for collecting information on the potentially fatal MVe), and so traps may
not necessarily be located in areas of highest RRv disease risk. In the RRv case study
areas there had been a long history of outbreaks, and the transmission cycle, vector
species etc., were well documented.

8.4.2 Reliable and up-to-date information on exposure and health outcome

Good disease data (historical and current) are needed to model the current disease risk,
and to predict future risk. "Preparedness" implies that potentially emerging events be
known (Richet et al. 2001). For this to occur, public health and disease-related
infrastructure (e.g. pathology laboratories) need the capacity for good clinical or
serological surveillance, and clear diagnostic criteria to confirm cases (Garner and
Kalunda 1999). Many industrialised countries have national disease surveillance systems,
and the reporting of infectious diseases is mandatory (Myers et al. 2000). Even so,
depending on the scale of the early warning system these data may not be sufficiently
accurate for prediction in all regions. In the case of RRv, a limitation of the notification
data (whereby cases are reported by postcode of residence, rather than postcode of
infection) dictated that modelling could not be reliably conducted in metropolitan areas.
The lack of consistency in diagnostic criteria between States for RRv disease has been
discussed previously. The majority of RRv notifications are believed to be presumptive
(Curran 1994, Curran 1996). In terms of an early warning system, diagnostic consistency
within a region would be important.

Meteorological data (e.g. rainfall, temperature, humidity, ENSO indices) are available at
various resolutions and quality, depending on the region and country. Access to data is
limited by, among other things, costs associated with its provision, limited availability in
terms of poor station coverage, and missing values (WHO 2001). In many countries,
remote sensing data provide a substitute for weather data when it is not available from
observation sites, and these are often available free of charge. The accuracy, scale and cost
of these data are a significant factor in the successful development of models. Additional
data, such as soil moisture proxies, or wind speed, may be desirable but are often not
available (as was the case in the study areas). For many mosquito-borne diseases, the
ecology of the pathogen varies widely between regions, and a suitable ecological zonal
classification is required to develop predictive models (e.g. see Garner and Kalunda
1999). In the case of RRv, GIS software was used to combine the major Australian climate and vegetation zones into regions within which conditions were assumed approximately similar.

8.4.3 A model that is accurate, specific and timely

Any formal attempt to predict the future occurrence of disease requires the use of modelling (McMichael et al. 2000). Chapter 5 has noted the types of models that can be employed for disease prediction. Not all predictive models would be suitable for use in an early warning system. From the experience provided by existing systems, the following section suggests the types of models that would be most useful.

8.4.3.1 Model accuracy

The accuracy of a model includes the elements of precision (consistent measurement, or reduced standard error) and validity (e.g. the number of epidemics/non-epidemics that are correctly identified) (Greenland and Rothman 1998). Predictions of disease epidemics can be made at several levels of precision (Garner and Kalunda 1999). As a rule, the value of greater precision in study results must be matched against the greater cost required to achieve it (Rothman and Greenland 1998). In the case of the RRv study area models, an increase in accuracy could have been achieved (i) with a longer time series, or (ii) if the scale of the unit of analysis was decreased (thus increasing the sample size). It is often the way, as it was in this study, that the case time series could not be extended (i.e., RRv notifications commenced in 1991, and data prior to this were non-continuous or unavailable). The issue of scale is discussed further below.

Highly accurate models may be neither possible nor warranted. For example, early warning systems can also make use of qualitative assessment criteria for decision-making. In instances where the availability of data and the understanding of the dynamics of transmission in a region is minimal, a simple scoring system could be employed to assess the potential for virus activity (based on general knowledge acquired from other regions) (Garner and Kalunda 1999). This approach would be relevant for RRv regions that are geographically close, or climatically similar, to other regions in which formal predictive modelling has been conducted.

Statistical modelling, by its nature, is more suited to regions that record high variability in the outcome variables (occurrence of epidemic). Thus, early warning systems for
mosquito-borne diseases are better suited to regions that record epidemic patterns of disease (with a strong inter-annual pattern of events) rather than endemic ones (in which a steady background rate of cases are reported, year in and year out, with slightly higher peaks in some years than others). These regions are easier to model than those with low variability, in which it would be hard to distinguish the epidemic “signal” from the background “noise”.

8.4.3.2 Timeliness and specificity

The development of an early warning system involves an inevitable compromise between the timeliness of the system (in terms of the lag between prediction and event) and the technical accuracy and reliability of the indicators. Medium-range predictions give the least specific warnings, but have the benefit of providing planners with long lead times (Myers et al. 2000). This is especially the case with the use of ENSO as a prediction variable: there are powerful advantages of a prediction three to eight months in advance, although the nature of ENSO teleconnections means there is no guarantee that rainfall (or lack of it) will necessarily arrive.

The WHO (HIMAL and MARA 1999, Thomson and Connor 2001) have proposed a useful framework for conceptualising the prediction stage of early warning systems. They distinguish three elements of “epidemic forecasting”, “early warning” and “early detection”. Epidemic forecasting predicts conditions in the future, usually using medium-range climate forecasts. It provides the longest lead-time, but will also be the least specific and reliable. Early warning involves the monitoring of local factors (such as weather, or mosquito activity) to provide greater reliability and local specificity, although this allows less time for response. The early detection stage includes monitoring of case data to provide highly specific and reliable information on where an outbreak is occurring. Obviously, at this point there is very reduced lead-time. Depending on the typical shape of the epidemic curve, however, action at this point may still be effective in reducing a substantial number of cases (or, for malaria, the severity of symptoms).

In terms of RRv disease, predictive models were developed for two stages (“early” and “late”). The timing of each was linked to the RRv transmission and infection cycle in the case study areas. This involved a trade-off: if the early warning period in the Murray was reduced by a month (from November to December) this could have increased the
accuracy of the prediction. It would also have reduced the effectiveness of response (i.e., vector control or public notification).

The timeliness of warnings relates to more than choosing an indicator that provides sufficient lag between the prediction and cases. It also relates to the "ability of the system to collect and analyse data quickly" (Huss-Ashmore 1997) and to the time taken to achieve a response. A warning one month prior to an outbreak is only useful if communication with relevant stakeholders and appropriate control measures can occur within that period.

8.4.3.3 Informal systems

Although the benefits of formally describing and forecasting epidemics are many, the role of existing informal early warning systems should not be overlooked (Buchanan-Smith 1997). In relation to RRv disease, the knowledge-base of some field entomologists and public health officers - who rely on their own observations of local events as a basis for decision-making, and who have established information sharing networks with colleagues from other regions - is considerable and highly sophisticated. Such local expertise should be incorporated into the development of an early warning system (in some regions these observations may be the only information available on the transmission cycle of RRv disease). A major disadvantage of an informal system of this type, though, is that it concentrates expertise among very few people, with the danger that the knowledge is lost from the region if the person leaves.

8.4.3.4 Appropriate scale of analysis

Buchanan-Smith (1997) categorises early warning systems into local, sub-national, national, regional, and global levels. In principle, the appropriate scale of an effective early warning system will be determined by the needs of the public health authority responsible for disease prevention, guided by consideration of (i) the cost-benefit of developing an early warning system at different levels, (ii) the complexity of the disease transmission dynamics, and (iii) the resolution at which disease and exposure data are available. The first of these points is discussed later in this chapter. In relation to the second, predictive modelling at the national or State level scale is not appropriate for RRv and many other mosquito-borne diseases, as already discussed. The findings from Chapters 6 and 7 demonstrate that biologically-meaningful analyses can be successfully conducted within RRv bioclimatic zones in which temperature, humidity, rainfall, and
vegetation are more uniform. Prediction at finer scales would also be biologically defensible, and may be practical for highly population dense regions.

The choice of the spatial unit (the SLA) was determined by the availability of population data, and consideration of the seriousness of the disease and end utility of the model. Investment in highly sophisticated systems is warranted only when the response actions are meaningful, the population is highly vulnerable, and the hazard is extreme (Committee on Climate 2001). For a disease with a high fatality rate and infectiousness (such as Ebola), extensive local analyses are justified to determine the mechanism of infection, and to attempt to generate forecasts of outbreaks. RRv disease is not such a high-priority disease: although there are on average some 5000 cases a year, it is non-fatal, patients typically recover without long-term sequelae, and treatment costs are small. Hence, the resource expenditure required for ongoing surveillance and the development of predictive models at the fine local level (i.e., 20-50 km²) would be disproportionate to the perceived public health burden for the regions studied.

8.5 DISEASE RESPONSE

A feasible response plan is an essential component of an early warning system, so that the benefit of early warning is realised (Committee on Climate 2001). Feasibility in this context relates to the ability to generate timely, effective, and affordable prevention activities. Unlike famine early warning systems, where the prediction developers and users are usually separate entities, for mosquito-borne disease early warning the same agencies (typically public health and local government agencies) are responsible for generating the warning and for providing the response.

8.5.1 A detailed response plan

8.5.1.1 Thresholds for action

Experience from famine early warning systems has demonstrated that decision-making takes time and can interrupt the prompt and effective mobilisation of resources in an emergency (Buchanan-Smith 1992, Saidy 1997). The process of turning the prediction of an epidemic into an operational response requires agreement in advance on the triggers for activity. A suitable prediction model will "quantify thresholds, sizes and durations of
deviations from normal” (Garner and Kalunda 1999) for the environmental and health indicators. Ideally, these levels should be linked to prevention and control plans (Thomson and Connor 2001) that are developed in advance of the epidemic, and preferably “concurrent with the development of the prediction component of the early warning system” (Foran and Brosnan 2000).

A sensitivity analysis (in this thesis “validation test”) can be used to examine how changes in data can influence policy decisions (McNeil and Pauker 1984). The interpretation of the probability of an epidemic depends on varying the criterion for what constitutes one. In this thesis, the probability threshold used to define an epidemic was 0.5 (in the range between 0 and 1). If a different threshold was used, the sensitivity and specificity of the predictions would change (demonstrated in the study area results). A higher (stricter) threshold would mean that more “true” epidemics would be missed (i.e., more false negatives would occur), but that the number of times the system called an epidemic when one did not occur (false positives) would be lower. The relation between the threshold, the false positive rate, and the sensitivity of the test is called a receiver operating characteristic (ROC) curve (Fletcher et al. 1988). ROC curves can be used to provide a framework for estimating the costs of disease prevention, and hence determining thresholds for action (discussed below).

8.5.1.2 Vulnerability assessment

For the purposes of RRv disease risk, vulnerability can be considered to have two principal components: (i) the risk or frequency of being exposed to a hazard, and (ii) the ability of an individual, group, or society to protect themselves from a hazard (Watts and Bohle 1993, Dilley 1997, Moseley and Logan 2001). Successful prevention requires attention to high-risk groups as well as high-risk areas. In relation to mosquito-borne diseases, some exposure risk will always remain (given that control techniques will never completely eradicate mosquitoes), and an attempt must be made to “increase the coping capacity and reduce the consequences of the exposure” (Dilley 1997).

The capacity of a population to adapt to an external stress (termed “adaptive capacity”) varies substantially from system to system, sector to sector, and region to region (IPCC 2001c). In relation to the threat of mosquito-borne disease transmission, it depends on the quality of public health infrastructure (Yohe and Ebi (In Press)), the availability and effectiveness of response options (i.e., vector control, prevention education campaigns,
ongoing evaluation), and the environmental context. Even in a country as relatively well-off as Australia, a community’s capacity to protect itself also depends on the availability of financial resources. Those that are already on the margins of coping will be least able to respond to changes in the system (Woodward et al. 1998). At the individual level, vulnerability to RRv is a function of a person’s age, their immune status, their exposure behaviours (e.g. house location, outdoor work), and their individual practice of personal protection measures (which depends on their perception that the risk is real and significant, the extent of behavioural change required, their financial resources, etc.). Thus, groups within populations will be at differential levels of risk. Outdoor workers were observed to have higher levels of RRv antibody in one study (Hawkes et al. 1985), perhaps due to the difficulty of repeatedly applying mosquito repellent in some working conditions, as well as to their increased exposure to mosquito bites.

Other early warning system experiences (Field 1991, HIMAL and MARA 1999) have noted the importance of locating, with reasonable specificity, the spatial occurrence of a hazardous event so as to enable effective intervention. The RRv models were developed within bioclimatic regions. Within each of these, risk probabilities were generated for each SLA. Thus, opportunities exist for the fine-tuning of surveillance and response down to the SLA level (e.g. mosquito trapping or vulnerability assessments).

Information about the groups at greatest risk of RRv transmission will improve the targeting of public notification campaigns. These data can be collected by epidemiological and sociological surveillance and research, or from collaborators in a region (i.e., community groups, unions, tourism boards). Surveys are useful for monitoring knowledge of personal protection measures, although they are more expensive and time-consuming than the routine collection of administrative data (Saidy 1997), and would be beyond the resources of most public health units.

8.5.2 Available and effective interventions

In general, the degree of sophistication of an early warning system should be matched to a country’s (or region’s) capacity to take meaningful action to the hazards being predicted (Committee on Climate 2001). For a disease like RRv, no treatments are available to mitigate the impact of disease, and prevention is the only response. Potentially, early warning of RRv disease epidemics should enable more sensitive timing of interventions. For example, education campaigns could be used only in “high risk”
years, to avoid fatigue from people currently exposed to the same message every year. Advance warning of an epidemic could also allow more effective and efficient vector control. Were these the case, an added benefit would be a reduction in the total amount of resources expended.

8.5.3 Assessment of cost-benefit and affordability

In public health, and in most areas of policy, prevention is better than cure for several reasons. Firstly, it seeks to avoid human suffering. Secondly, the approach maximises the use of limited public health resources. Even so, prevention costs money. RRv affected public health regions might support the establishment of an early warning system in theory. The realities of scarce resources and competing priorities for the health dollar would demand that an assessment of the systems costs and benefits be conducted before proceeding. Such an assessment would determine whether a system was worth implementing at all, and if it was the most efficient way of achieving the greatest possible social benefit (Hutubessy et al. 2001). Furthermore, an early warning system must also be affordable, as “any intervention can be undertaken only if it takes resources from some other area” (health or non-health) (Hutubessy et al. 2001).

In general, the benefit of preventing an additional case of disease (the marginal benefit) should be equal to or greater than the cost of preventing a case of disease (the marginal cost) (Last 2001), although this obviously depends on the political nature of the disease, and the stage of its establishment in a community or region. For example, the prevention costs for West Nile virus disease in the first three years following identification of the virus in the USA in all likelihood vastly outweighed the costs associated with the cases that were identified during that period. Nonetheless, it was considered socially desirable to attempt to avert a far larger number of future cases, and this uneven expenditure was considered to be justified under the circumstances. The relatively low morbidity of RRv disease, combined with the fact that it has been long established in the Australian region and is one of only many competing health priorities, means the costs of intervention would definitely need to be less than the cost of the status quo. Depending on the sophistication of the system required, and the seriousness of the disease, monetary and non-monetary benefits may both need to be quantified, or at least stated if their nature is unquantifiable (Martin et al. 1987).
Detailing the annual costs averted by an RRv early warning system would be essential for its ongoing survival, as there is a general problem with maintaining government interest in funding public health prevention activities. Two examples are relevant. The cessation of the historic malaria epidemic forecasting system in the Punjab (which followed the partitioning of India in 1947) appears to have resulted, in part, from the system’s success. A decline in enthusiasm for implementing the system had been apparent from the late 1930s, by which time the “dreadful fever epidemics that had motivated the system’s establishment” faded from public health workers’ memories, and the maintenance of the system became a low priority (Butler 1997). In Australia, the funding for continued HIV/AIDS education of sex workers has been constantly under threat since the mid-1990s, given the extraordinarily low HIV transmission rates recorded between workers and their clients. Sex worker groups attribute this Australian anomaly (rates in neighbouring Asian countries are very much higher) to the effectiveness of continuous (funded) HIV prevention education, combined with law reform.

An estimation of the cost of an early warning system also needs to count the cost of a type I (false positive prediction) or type II (false negative prediction) error. For example, even if the system has a high sensitivity (i.e., the number of epidemics correctly predicted is high), the number of “false alarms” may be too high. If this is the case, the cost of attending to these may outweigh the overall benefit that the real alerts provide.

8.5.3.1 The cost of RRv disease

Although RRv disease affects a considerable number of people in Australia each year, a comprehensive assessment of the costs of the disease to the community has yet to be prepared. As is often the case, many data needed for this estimate are unavailable, available only for very small population groups or regions (and therefore not necessarily generalisable), or are unvalidated. In addition, even if accurate costs were available, it is likely they would vary between regions, depending on the method of vector control used.

A number of monetary costs have been estimated. Harley and others (2001) provide the most detailed and up to date estimate to date of annual disease costs, summarised in Table 8.1. They calculated the annual total costs incurred by all cases during the period 1991 to 2000 to lie somewhere between A$3.1 and $6 million per year (one Australian
dollar is approximately equal to US$0.6). The average cost per case (derived from these figures) was between $650 and $1265.

Table 8.1  Partial estimates of the annual cost of RRv disease in Australia, for the period 1991-2000.

<table>
<thead>
<tr>
<th>Type of cost</th>
<th>Cost category and details</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases</td>
<td>• Average number of cases per year (1991-2000) = 4745</td>
<td></td>
</tr>
</tbody>
</table>
| Testing          | • Pathology cost for single RRv test = A$38.40  
• Percentage of positive tests (5%-25% of tests ordered) 
• Total number of tests ordered per annum = 18,980-94,900  
**Total cost of testing (positive and negative)** A$728,832-$3,644,160  
|                  | **Total cost of medical care** A$217,796  
• Cost of two consultations with family doctor = A$65.90 |
| Lost earnings    | • Average time off work = one week in 50% of cases 
• Average Australian weekly earnings before tax = A$908.20  
**Total lost earnings** A$2,154,704 |
|                  | **Total estimated costs** A$3.1m to $6m /year |

Table based on analyses published by Harley and others (2001), updated as noted below.

- This cost includes the cost of the test (IgG or IgM, both of which are $23.80) and the cost of a collection fee ($14.60).
- Does not include cost of testing to exclude differential diagnoses.
- Does not include cost of medications.

These figures only provide a partial estimate of the total costs, however, as they are based on data for one State of Australia (Queensland). The pathology costs are likely to be underestimated: they do not include the cost of testing to exclude differential diagnoses, and only allow a single test to confirm RRv status (two tests are sometimes conducted to confirm a positive diagnosis). Furthermore, the assessment does not account for treatment costs, patient travel to and from a general practitioner, or decreased work output of more than one week. Even if RRv disease is self-limiting for the majority of patients, and symptoms progressively resolve over three to six months (Harley et al. 2002, Mylonas et al. 2002), there has not been any estimation of the costs of physical pain,
family support, etc. during this period. The disability-adjusted life year (DALY) measure could be a useful method to employ in this regard (Murray and Lopez 1996).

In addition to these costs, a comprehensive assessment should account for the costs incurred by State and Local Governments in disease prevention, and to the environment from the negative impacts of control measures. Examples include the cost of vector control, education campaigns, and research, and the possible costs of drug resistance or ecological damage.

8.5.4 Involvement of all partners in the system

Thomson and Connor (2001) have noted the limited routine use of early warning systems in Africa for malaria control programs due, they argue, to poor inter-sectoral collaboration between health and the meteorological and agricultural sectors, as well as to the lack of evidence of the cost-effectiveness of these systems. For a disease like RRv, the primary stakeholders (i.e., those involved in the operation of the system) might include public health regional units (responsible for disease surveillance and public education), Local Government authorities (responsible for vector control interventions), and staff in the meteorological organisation that provide regular weather data. Other people who might need to be involved in establishing a system could include staff of State Government Health Departments who can direct funds or provide non-financial support, and researchers, statisticians, entomologists, and others with expertise in RRv issues.

8.5.5 Communication strategy

Even reliable and specific warnings will fail to achieve their objective if the message is inaccurately or improperly transmitted (Tapscott 1997). Walker (1989) identified four principal components of a successful early warning system, two of which relate to the transmission of information. These can be summarised as:

1. constructing the warning or response message (in lay terms, detailing what is predicted, and who is most at risk, in a consistent manner), and
2. distributing the message (accurately, and through the correct channels).

The RRv models were developed at the regional level, but monitoring and response for an early warning system would be more likely to be conducted by clusters of Local Government shires within the regions. In a developed country such as Australia,
channels of communication are technically excellent. However, good communication consists of more than just technical capacity. In the context of plague locust early warning, Foran and Brosnan (2000) have recommended the development of a “chain of communication” as part of a response plan. This details the people involved from the point of prediction analysis, through to those responsible for transmitting, receiving, and responding to the warning. It should also indicate the maximum time for communication between collaborators. Even if communication between the main parties involved in the system is effective, experience from other systems (e.g. HIMAL and MARA 1999) has demonstrated the importance of maintaining communication with peripheral agencies. These could include the media, community organisations, health committees or school bodies involved in distributing education messages, etc.

In the initial years of a system, achieving consistency of response messages between collaborators in a region may be a simple matter. As time goes on, however, and with additional research into adaptation strategies, divergent views on how best to promote disease prevention (at the public education campaign level, as well as at the individual protection level) may emerge. This occurred between Australian States in relation to HIV/AIDS education messages in the 1990s, and the sometimes contradictory information that ensued reduced public confidence in the self-protection measures that were recommended.

8.6 THE CONTEXT FOR RRv EARLY WARNING IN AUSTRALIA

8.6.1 Current RRv disease management

Disease control aims to decrease the incidence or prevalence of a disease, or to eliminate the conditions that contribute to outbreaks. No vaccine is available to prevent RRv infection. Therefore the principal disease management activities can be categorised into (i) surveillance (of the vector, pathogen, or human populations), (ii) source reduction and control of vectors, and (iii) public notification of risk. RRv disease management is a joint responsibility of both Health Departments and Local Government authorities in all States and Territories, with different jurisdictions adopting different models to apportion costs
and responsibility for surveillance, control and prevention activities.\footnote{In addition to the sources cited in the text, information on current RRv disease management in Australia was collected directly from State and Territory Health Departments. I requested information from all health departments of their current RRv management activities, including copies of arbovirus reports, strategies, and other relevant material. This information was further supplemented by follow-up phone calls to some Health Departments and Local Councils.}

8.6.1.1 Surveillance

Disease surveillance is used to assess public health status and to define priorities. It involves the systematic ongoing collection and analysis of data, and the timely dissemination of information so that action can be taken (Last 2001). Surveillance data indicate who is affected, in what geographic area problems are occurring, and where interventions should be directed. The surveillance mechanisms for RRv disease include case data collection, surveillance of vector and virus activity, and meteorological surveillance.

Primary surveillance for RRv disease involves the collection of RRv notifications, case follow-up data, or serosurveys to determine population immunity levels. All States and Territories in Australia record RRv notifications from laboratory serological tests, and mandatory reporting to the National Notifiable Disease Surveillance System is required. Questioning of cases to determine the likely place of infection is practised routinely only in Western Australia and South Australia, with a varying response rate. Some other States occasionally conduct follow-up of cases at the beginning of the season to determine high-risk spots for mosquito breeding. No routine serological surveys are conducted in Australia for RRv, although several serological studies have been undertaken in the past (e.g. Boughton et al. 1984, Fraser et al. 1986, Hawkes et al. 1993). Testing of human sera from blood banks for RRv-specific IgM has been proposed as an RRv surveillance method by Weinstein and others (1994), although there is no indication that this is being considered as a viable option in any Australian State.

Mosquito and virus surveillance programs are undertaken to a limited extent in some States (Northern Territory, Western Australia, New South Wales, Queensland, and Victoria). In New South Wales, for example, mosquito populations are monitored in areas with large human populations or a long history of arbovirus disease (NSW Department of Health 1998). The initial purpose of this mosquito surveillance program was to monitor flaviviruses only (specifically MVe: NSW Department of Health 1998), and traps were positioned according to high risk regions for MVe activity. The program was later
expanding to include alphavirus surveillance (Ross River and Barmah Forest viruses), which suggests that the siting of traps may not necessarily be optimally positioned for monitoring RRv activity. As noted previously, south-west Western Australia also conducts regular mosquito surveillance. A surveillance scheme utilising the sera of possible vertebrate hosts such as macropods, sheep, pigs, and rabbits has been suggested (Kay and Aaskov 1989). However, this has not yet been considered financially viable.

Meteorological surveillance now forms part of the arboviral surveillance programs in several States. For example, rainfall, temperature and tidal data are collected in parts of south-west Western Australia (e.g. Broom et al. 2001) and New South Wales (NSW Arbovirus Surveillance & Vector Monitoring Program 2003b). In Western Australia this information (in particular tide height) is principally used to dictate the timing of mosquito trapping runs, and to describe variations in mosquito density and cases for reporting purposes. Neither State formally use these data to provide advance warning of the likelihood of epidemics of RRv disease.

8.6.1.2 Vector control

Traditional reliance on adulticiding (the spraying of adult mosquitoes) for the control of mosquitoes has been replaced, in Australia as in many other countries, by an “integrated pest management” approach to minimising the risk of disease (Rose 2001). This involves the use of biological control pathogens, such as predators, in conjunction with “biorational” pesticides (Mulla 1994). In addition, the modification of mosquito breeding sites by environmental engineering (e.g. filling in swamps, runnelling, water management, removing vegetation near human populations) is also conducted in some regions.

In most States, mosquito control is primarily the responsibility of local governments. Ground or aerial spraying of larvae is practised in only a few States (Western Australia, Northern Territory, Queensland and, on occasion, Victoria). The largest mosquito control program in Australia is conducted in Brisbane, Queensland, a city that is surrounded by salt marshes (some 30 000 hectares of salt marsh mosquito breeding habitat are aerially treated each year). Adulticiding is undertaken rarely in some States, but is not practised at all in most, as it is generally considered ineffective. In New South Wales and South Australia (States where the Murray study area was located) there is a strong focus on personal responsibility for protection rather than a reliance on mosquito control (NSW
Department of Health 1998). With the exception of a few local government projects in the areas of Tweed Heads, Port Stephens and Griffith, mosquito spraying is not routinely practised in New South Wales (NSW Department of Health 1995).

Source reduction involves the elimination of larval habitats (Rose 2001). Runnelling, a form of habitat modification, allows water to drain from trap pools along natural drainage lines to connect with tidal sources (Russell 1998a), and permits predatory fish to gain access to mosquito larvae. Runnelling has been successfully used in parts of Australia, although it is not suitable for regions that generate large, flat fields of water (such as around Darwin, Northern Territory, or in the Murray River study area). A variety of larvivorous predator fish and parasites have been proposed for the biological control of RRv vector species in New South Wales (NSW Department of Health and ICPMR 2002).

The design and siting of settlements can also influence human contact with mosquito breeding sites. Tai and others (1993) found different attack rates between two towns in the Northern Territory, even though both areas had identical rainfall patterns. They attributed the lower incidence of infection in one town to the groundwater control systems, which drained water away from the ground surface. In the Northern Territory, all development planning applications (i.e., residential estates, dams) are obliged to be considered by the Health Department for comment. In New South Wales, land planning legislation at the local and State level may require a developer to take into consideration the possible impact of a proposal on mosquito breeding, as well as the proximity of dwellings to existing breeding sites (NSW Department of Health 1998).

8.6.1.3 Public notification and education

Community education to reduce exposure to mosquito bites is important, given that vector control will usually be inadequate. Larviciding will never completely knock out mosquito populations in years of heavy rainfall or high tides (Rose 2001), and is economically impractical in rural areas where human population density is low.

Primary prevention of RRv disease involves action taken by the individual or community to reduce the risk of being bitten by mosquitoes. Primordial prevention, on the other hand, consists of public health policy measures that seek to inhibit the environmental,
economic, social, and behavioural conditions that are known to increase the risk of disease (Beaglehole et al. 1993). Such activities include workplace education for workers who are regularly exposed to mosquitoes, or the use of planning instruments to discourage or reduce development around mosquito-breeding sites.

In all States, prevention education is the responsibility of the State Health Departments, occasionally in concert with Local Governments. The approach varies slightly, depending on the climate of the region. In the tropical north, “low level” education all year round is combined with media alerts when an increase in cases is detected. In the temperate south, education generally only commences as the weather warms at the start of summer, and when mosquito activity is more likely (or, in those areas that have the resources for mosquito surveillance, when activity has been actually observed). States adopt a variety of similar education techniques, including the use of written material (such as pamphlets explaining the recommended protective measures, placement of warning signs in high-risk recreational areas) and health alerts via the media when high numbers of cases have been recorded. Both local residents and tourists can be targeted, again depending on resources and region.

8.6.2 How could the current approach be improved?

8.6.2.1 Ineffective prevention education

There is accumulated knowledge through international research of other diseases (such as malaria and dengue), and through some Australian research (Weinstein and Cameron 1991, Murray-Smith et al. 1996, Westley-Wise et al. 1996), that the wearing of long-sleeved loose clothing, use of personal mosquito repellent, insect proofing of buildings and water tanks, and removal of domestic breeding sites, will reduce infection rates for some mosquito-borne diseases. No studies in Australia have investigated the comparative effectiveness of these different personal protection strategies, however. Research from other countries on mosquito-borne diseases often cannot be generalised to the experience of RV disease Australia, given differences in vector biology and pathogen transmission, or for social or cultural reasons, etc.

Arbovirus prevention campaigns generally lack the socio-behavioural information needed to motivate people in various social settings (Knudsen and Sloof 1992). Research
into the factors that increase or decrease the use of personal protection, and the appropriate targeting of information in Australia is scarce. More information is needed, for example, on how knowledge of local vector control activities affects people's perception of their risk of infection. Do they become less attentive to self-protection measures? How often should prevention messages be delivered over the course of a season, and by whom? Is it better to provide specific details about the habits of mosquito vectors in each region, or will general (and basic) information be more effective? Should particular groups in communities be targeted, or would a blanket population approach be better? Assertions about the merit of particular self-protection measures (such as modifying behaviour to avoid peak biting times or areas of high mosquito abundance) have not been accompanied by research that addresses the substantial obstacles to the practice of these measures.

Given the lack of information noted above, regional public health authorities are forced to rely on simplistic prevention messages (NSW Department of Health 1998) for what are quite complex behavioural changes. Given minimal resources in most regions in Australia, education campaigns are almost never evaluated.

To improve the quality of data, differences in risk perception between regions and groups would need to be identified. In rural communities, for example, it has been noted that there "is an attitude of apathy or stoicism" to contracting RRv disease (Russell 1998a). If true, this would obviously affect the level of prevention education uptake: annual repetition of prevention messages may dull people's minds and cause them to "switch off" to a risk they perceive to be constant.

8.6.2.2 Economic, environmental, and social sustainability

The use of vector control strategies for RRv disease will continue to be appropriate in certain ecologies and at particular times. As a long-term public health strategy to reduce disease, however, mosquito control has several problems. Weinstein (1997) has argued that the use of insecticide sprays to control mosquitoes is appropriate at the interface between endemic and non-endemic areas, but not in places where non-immune populations enter an endemic area (e.g. as a result of urban or agricultural expansion, or tourism). This is because (i) endemic areas are likely to be large, and spraying is not cost-effective in the long-term, and (ii) it will delay the exposure of children to RRv until they
are likely to show symptomatic disease. The use of insecticides for mosquito control in these environments can lead to an increase in the incidence of disease. Instead, education about anti-mosquito measures would be more appropriate (Weinstein 1997).

There is steady opposition to the use of insecticide control, for both health and environmental reasons. Although only used at low rates, many insecticides are toxic to insects, birds, fish, bees, and aquatic life (Rose 2001). Larviciding is not permitted in fish habitat areas or conservation parks in Queensland (Queensland Department of Health 2002). Even though the area recorded some of the highest rates of RRv disease during an outbreak in 1984 (1070 per 100 000 people: Hawkes et al. 1985), the Griffith Council did not use insecticides to control mosquito activity due to local community and industry concern about the flow-on effect to consumer confidence of fruit grown in the area.

Furthermore, applying insecticides without complete eradication of the mosquito population can increase the chance of building resistance (Weinstein 1997). Increasingly, efforts are being directed towards the development of methods for mosquito-borne disease management that are environmentally, economically, and socially sustainable. Such approaches include disease control plans that are developed in consultation with communities (Knudsen and Sloof 1992) and the management of pesticide resistance (Mulla 1994). Economies of scale enable resource-intensive source reduction in more densely populated regions, such as alterations to the hydrology of marsh ecosystems (Hulsman et al. 1989) and sub-soil drainage (Whelan 1989). However, these methods are not effective in controlling mosquito numbers and minimising environmental impacts in all areas (Dale et al. 1989).

The economic, social and environmental issues of RRv disease management need to be addressed simultaneously. Sustainable solutions require that all stakeholders (government, community, and industry) determine the "net benefit" of preventive actions for future, as well as current, generations (Government of Western Australia 2002). Because of the requirement for community involvement and attention to local issues, solutions will necessarily be locally specific and may not be directly generalisable to other regions. The processes, however, might be.
8.6.2.3 Tourism in Australia

Tourism is one of Australia’s fastest growing and most significant industries: domestic and international tourism accounted for 4.5% of the Gross Domestic Product in 1997-98 (Commonwealth Department of Industry, Sport and Tourism 2002). More than 4.8 million people from overseas visited Australia in 2002, 55% of them for holiday purposes. International visitor arrivals are predicted to increase to 8.1 million by 2012, an average annual increase of 4.8% (Tourism Forecasting Council 2002). The growth of tourism into northern tropical and sub-tropical Australia, predominantly to non-urban areas, suggests the risk of infection among non-immune visitors will increase (Kelly-Hope et al. 2002). A mechanism to provide advance warning of outbreaks at the SLA level could help with the targeting of information towards tourists.

8.6.2.4 Population pressures

In addition to improving the current situation, future population growth and human settlement patterns are likely to place increasing numbers of people at risk of exposure to RRv disease. The Australian population is expected to grow from 19 million to between 25 and 28 million by 2051 (Trewin 2000). This represents an increase of 9-38% in the total number of people in the RRv higher risk age-groups (i.e., 15 to 65 years of age). Increasing urbanisation has been accompanied by a general trend from rural to city living in Australia, as elsewhere. Currently about 86% of the Australian population live in urban areas. This is expected to rise to 96% by 2050 (Trewin 2000). The current trend (predicted to continue) is that growth will move outwards from existing urban areas, rather than increasing population densities within current city limits. Between 1986 and 1991, 28% of Australian population growth occurred in non-metropolitan coastal regions (State of the Environment Advisory Council 1996b), mostly into productive farm land and bushland. If this future picture holds true, non-immune humans are progressively likely to be exposed to the “natural cycle” of the virus between the mosquito and reservoir hosts. In regions where the virus is enzootic, this presents a significant public health management issue. Because of these changes, the impetus for more precise methods of determining when and where outbreaks of disease will occur is also likely to increase.

17 Large outbreaks of RRv disease have followed residential developments in the previously unsettled salt marsh habitats of the Peel and Leschenault regions in Western Australia (see Mackenzie et al. 1994b, Russell 1995).
8.6.3 Summary

Increasingly, greater attention is being paid to sustainable solutions for social, environmental, and economic problems. Each Australian State Government and many Local Governments are in various stages of developing a Sustainability Strategy to set the long-term direction for their regions (Newman 2001). The current approach to RRv risk management in Australia, although following the world's best practice in terms of integrated pest management, has a number of intractable problems. Concern for possible ecological and human health impacts, the possibility of vector resistance, and the increasing financial costs of vector control, all favour limiting the use of sprays and environmental modifications as much as possible.

Some of the elements needed for an early warning system exist in certain regions of Australia (i.e., meteorological and disease surveillance, strong collaboration between stakeholders, a knowledge of the RRv transmission cycle). There is, however, no mechanism for discriminating between "high-risk", "medium-risk", or "low-risk" seasons. Accurate and timely prediction, coupled with timely and effective response, could limit vector control and public education to those years when there is a high risk of epidemics. Hopefully, this would have the benefit of reducing cases, reducing chemical load on the environment, and stretching limited resources further.

8.7 A hypothetical RRv early warning system

8.7.1 Overview of the early warning system

The following section provides an overview of the stages of a hypothetical RRv disease early warning system. Each stage is associated with specific indicators and responses, and the amount of warning time and accuracy are noted. For simplicity, the details are based on the findings from the Murray study area only. Figure 8.1 is a diagram of the proposed three-stage early warning system.
8.7.1.1 Stage One (epidemic forecasting)

**Indicator**
- Less than three days of rainfall recorded during the austral spring to early summer (October-December) of the preceding year. The probability of vertebrate host immunity is decreased (and hence the potential for virus amplification in the following year).

**Warning Time**
- This forecast would precede the start of an epidemic by some 11 months.

**Accuracy**
- Sensitivity = 88%, Specificity = 47%.

**Response**
- A low number of rain days would trigger a first-stage “alert” in relevant SLAs. Specific activities might include an increased commitment to resources for mosquito surveillance in the following spring, attention to maintaining spray equipment, checks of communication linkages and procedures, and notice to be ready for mosquito control actions in the following year, depending on the result of Stage Two.

8.7.1.2 Stage Two (early warning)

**Indicator**
- Climate data (as per early warning model in Chapter 6) would be analysed for the months of July to November. From these data, the model would generate the probability of an epidemic for each SLA.

**Warning Time**
- From early December (two to four weeks notice prior to the optimal period for larviciding). Initial RRv disease cases commence in November and December (peak cases are not recorded until March or April). This allows between several weeks and four months for staggered public education campaigns (and three weeks until the start of the main Christmas holiday period). Risk groups could include fruit pickers, tourists, etc.

**Accuracy**
- Region 1: Sensitivity = 62%, Specificity = 95%. Region 2: Sensitivity = 73%, Specificity = 81%.

**Response**
- Larviciding and public education (noted above). If resources allowed, monitoring of mosquito populations and virus activity could be conducted during November in SLAs identified as high risk in Stage One.
Stage Three (detection)

**Indicator**
- Climate data (as per late warning model in Chapter 6) analysed for July to February. From these data, the model would generate the probability of an epidemic for each SLA. The record of RRv disease cases notified to that point (i.e., early March) provides additional certainty and some indication of the likely magnitude of the outbreak.

**Warning Time**
- At this point vector control would not be effective. However, on average only 54-58% of cases are reported by the end of February. Additional education campaigns could still be effective in reducing some 40% of cases.

**Accuracy**
- Region 1: Sensitivity = 96%, Specificity = 93%. Region 2: Sensitivity = 66%, Specificity = 98%.

**Response**
- Additional surety about the probability of an epidemic provides the opportunity to focus prevention education messages to those most at risk.

The suggested interventions outlined above are triggered by pre-defined thresholds. Some SLAs in a study area might choose different cut points to define the probability of an epidemic (these examples used 0.5), depending on the result of cost-benefit analyses, a local risk aversion assessment, etc. Similarly, if surveillance was conducted, mosquito trap density, numbers of virus isolated, and the number of human cases that constituted a “threshold” would need to be quantified. It has been observed that early season population reduction for RRv vectors is more effective than spraying during an outbreak (NSW Department of Health 1997). Quantifying “early” and “late” in this context would need to be made.

8.7.1.4 Inputs required

The forecast support tools that would be needed for this type of ongoing prediction are:
- SLA digital boundary file.
- GIS and statistical software packages.
- Interpolated monthly summary values for the climate variables for each SLA, available online (Queensland Department of Natural Resources and Mines 2002). The Internet provides access to data in near real-time, when only a brief time ago reports of weather events were sent by mail, with several weeks delay. Data collection
This hypothetical system is based on the models developed for the Murray case study area. The system has three warning stages. The indicator for Stage 1 measures the number of rain days for the spring to early summer of the previous year. Low rainfall is linked to decreased mosquito activity, and thus higher vertebrate host immunity in the following year (one precondition for an epidemic). An early warning would be issued if the number of rain days fell below a threshold. In Stage 2, rainfall, temperature, and sea-surface temperatures are monitored for the winter and spring months of the current year (austral July to November). If the probability of an epidemic exceeds a defined threshold, a second warning would be issued. In Stage 3, summer rainfall and temperatures are monitored, along with RRv disease cases (which commence at the beginning of summer, and peak in autumn). If the threshold was exceeded, an epidemic would be confirmed at the beginning of March (the start of the austral autumn). Additional public awareness at this point could still be effective in reducing the total number of cases.
depends on regular (monthly) internet FTP data feeds of modelled climate data as input for the predictive models. Although available for a reduced cost for bulk purchasers, the climate data would need to be collected for eight months of each year, and represent the largest cost for the forecast end of the system.

The system inputs could be located centrally, with several people trained to keep the system operating (although only one would be needed at any time). Tasks could involve the regular collection, incorporation, analysis, and reporting of model outputs to relevant stakeholders.

8.7.2 General constraints on early warning systems

Theoretically, advance warning of an epidemic gives health authorities enough time to conduct mosquito control and to provide warnings to vulnerable groups. The reality is that success is not guaranteed. Due to a combination of political and institutional factors, as well as logistical and resourcing ones, there is often a failure to capitalise on an early warning (Buchanan-Smith 1992). In the case of RRv vector control, staff need to be regularly trained to ensure high technical accuracy in the application of chemicals. There is little tolerance with the timing of spraying in relation to the breeding season and local weather conditions. Equipment failure, absent employees (Rose 2001), and lack of chemical availability and aircraft or ground vehicles when needed, are all factors that could compromise the effectiveness of larviciding, and that would need to be managed to get maximum benefit from a forecast.

The major political budgetary cycle that operates at the State Government level presents a serious constraint to fully exploiting the resource efficiencies of advance warning. Financial management within government portfolios demands that the annual allocation of monies for programs be spent each year, and any surplus revoked. Advance warning should enable mosquito control savings in non-epidemic years to build a substantial amount of reserve for large-scale control in epidemic years. Instead, mosquito control activities must be conducted at a moderate level in all years to maintain an ongoing budget (Michael Lindsay pers. comm. 2002). Local Governments in Western Australia (not subject to the same budgetary limitations as the Western Australian Government) have commenced a system of retaining 10% of their mosquito control budget in non-epidemic years. This allows for approximately twice as much larviciding activity in
epidemic seasons to “normal” ones (every three to four years on average: Michael Lindsay, pers. comm. 2002).

A statistical model that predicts future patterns of disease is limited by a reliance on past or current data, and these relationships will not necessarily remain the same in the future. Even if all known confounders were incorporated as variables in the model, future changes in climate, environmental conditions, and social factors will be likely to influence the RRv disease transmission cycle. Future climate changes predicted for Australia make it likely that RRv activity will decrease in some parts of the country, and increase in others, with the pattern of change not yet clear, given the range of viruses and different vectors in different circumstances and regions (Russell 1998b). While the specific regional impacts remain to be researched, it is reasonable to assume that long-term climate changes may limit the predictive power of current observations at future time points (Committee on Climate 2001). In addition, if our climate does become more suitable for the transmission of RRv disease, human adaptation (especially public health interventions) could result in improved methods for combating transmission, and hence may modify the effect of environmental and climatic exposures.

8.7.3 The benefits of early warning

As noted in Chapter 1, climate forecasts will be helpful for the public health management of RRv disease if they can (i) improve the sensitivity of surveillance and increase the length of the response time above existing surveillance activities, or (ii) reduce the cost of traditional surveillance activities. The review of current risk management activities presented in this chapter, and the findings from the RRv disease models, suggest that in most regions prediction based on climate data alone would be an improvement on the accuracy and timeliness of current approaches.

With forethought of the end use of prediction models, this study intentionally used data, software and statistical techniques that are currently available to public health authorities, or that could be obtained with minimum effort and expenditure. Complex modelling techniques, and climate and environmental data that are hard to access (e.g. soil moisture data) were deliberately avoided. The interpolated weather surface, generated daily for the whole of Australia from current observations, is available via the World Wide Web. Similarly, the choice of scale for the development of models reflects the limited resources for RRv risk management. The models were developed at the regional
level, but the opportunity exists to increase predictive accuracy by incorporating more
detailed information about mosquito density and epidemiological surveillance at the SLA
level.

The benefits of the proposed RRv system have yet to be formally tested. Theoretically,
advance warning of the likelihood of epidemics enables (i) better communication of risk
and more sophisticated and specific prevention information, (ii) improved timing of
insecticide spraying, (iii) fewer cases of disease, and (iv) more efficient use of public
health resources.

Although RRv disease vaccination is not currently available, research is being conducted
into the development of a vaccine (Yu and Aaskov 1994, Aaskov et al. 1997), and human
trials are being considered. The RRv prediction models could potentially be used to
inform decision-making for an immunisation schedule. For example, output of the model
could categorise local areas into high (>70% probability), medium (30-70%), or low
(<30%) risk each season. This would enable people with chronic rheumatic conditions
living in the regions, or tourists travelling to the area, to consider the merits of
vaccination.

8.8 Conclusions

This chapter has argued the value of early warning systems as a planning tool to increase
preparedness of RRv disease in parts of Australia. Excellent predictive ability is of little
value, however, if the response capacity is not equally as good. Response can be impeded
by political, institutional, and administrative factors. In fact, the major work of
establishing an early warning system is ensuring that an advance warning can be
smoothly responded to. As well as developing models for prediction, public health
capacity would be well-served if future research and policy also focused on the less
technologically sophisticated (but no less important) details of action thresholds,
vulnerability assessments, communication strategies, etc.

The lack of information about the future impact of climate change on the ecology of RRv
is a problem in terms of devising adaptation policy and measures. At present it is unclear
how the incidence of RRv disease will change. Continued monitoring of the vectors in
regions that are currently epidemic - and particularly at the borders of these regions -
could be useful for building evidence of changes in species type, distribution, or breeding patterns (information that could be used to adjust predictive models). The resources available for RRv risk management are potentially high in Australia, but the current distribution of funds is low, as RRv is not perceived to be an especially high public health priority at the national or State levels. This limits the effort that can be directed towards developing adaptive capacity.

Social institutions and governance are comparatively accessible and equitably distributed in Australia. The relevant State public health and Local Government institutions are well established and delineated, and provide an institutional capacity to respond to future changes in RRv disease risk. The human capacity to respond, however, is unbalanced. Access to and use of technology for the management of mosquito-borne diseases is high by world standards. Even so, the current technology is inadequate as regards effectiveness and ecological sustainability. Although there is openness to new developments in this area, a quantum shift in the approach to mosquito control is not foreseeable in the near to medium term (Mulla 1994). While there is a high level of expertise in the RRv affected regions of Australia regarding mosquito control, the knowledge of the factors that support people making positive health behaviour changes to prevent mosquito-borne disease infection is poor. There is a lack of research in this field, and a lack of personnel skilled in preventive education strategies that have been tested in Australian populations.

If effectively designed and implemented, early warning of suitable climate conditions for outbreaks of RRv disease can provide the opportunity to prevent cases, and reduce public health expenditure. By using climate data in a staged prediction model (i.e., where forecasts increase in specificity as the time prior to the commencement of the epidemic season decreases), and combining them with prospective surveillance and infection rates of vector populations (if available), disease control activities could be improved by increasing the length of the response time. This could allow for more accurate communication of risk, and improved and more cost-efficient public education and vector control (i.e., activities could be limited to those years when there is a high risk of epidemics).
Chapter 9

Conclusions and further research

There is great potential for the use of climate and environmental data to provide early warning of RRv disease in Australia. Findings from this thesis suggest further research in a number of areas.

On the ecology of the virus

Hughes and others (2003) made the point that all mathematical models are necessarily a simplification of reality. This is true with models of the influence of climate on mosquito-borne diseases. Although the ecology of RRv has been substantially studied over the past 40 years, there is still a lack of knowledge about basic parameters with respect to vector breeding, the involvement of vectors and vertebrates in the transmission cycle in different regions, and the influence of climate on vector breeding (Mackenzie et al. 2000). For example:

• Observations suggest that the major RRv vectors in south-west Western Australia have remained consistent over the recent past. Even so, questions remain about the importance of *Ae. vigilax* vs *Ae. camptorynchus* as the major transmission vector in the summer months. In the Murray River area, the situation is more complex. Reports of RRv vectors have varied, and several new amplification and transmission vectors have been implicated in the recent past. High quality, up-to-date mosquito data requires a very large commitment of resources from both the entomological and public health communities. The lack of such information is likely to remain an ever-present issue for RRv disease modelling.

• Information about the influence of specific weather conditions (especially temperature, but also humidity) on the breeding rate and survival of RRv vector
mosquitoes, and the amplification rate of RRv, is sparse. This placed obvious limitations on the types of hypotheses that could be examined in this study. For example, more information on the effect of temperature on the biology of RRv species would have enabled specific temperature thresholds to be tested.

Research is an iterative process of building knowledge and generating questions. As additional data improve our understanding of the mosquito-climate relationship, changes to the environment may outdate that knowledge. A major difference in the future is likely to be the shortened length of this timeframe due to the expected rapid pace of climate change, whereby relationships may only hold true for short periods.

*On spatial analytic methods suitable for infectious disease transmission*

Regarding the details of modelling the climate and mosquito-borne disease relationship, several points are suggested. First, the method for selecting the boundaries for analysis (i.e., defining the RRv bioclimatic regions) was based on a composite of long-term average climate zones. It would be ideal if the classes within the climate and vegetation zones could be made more relevant to mosquito-borne disease research. For example, the breeding and survival thresholds for the major RRv vector species could be used to define the classes of the temperature zones. Second, statistical methods that allow for better management of dependence in the outcome variable could improve the modelling of epidemics.

Given the variations between regions in the component causes of RRv disease outbreaks, the modelling was a balance between generalisability and precision. For greatest accuracy, predictive models were conducted at the bioclimatic zone level – even though this meant the findings cannot be directly used in other zones. An extension of this work would involve repeating the modelling zone by zone.

On a more general level, modelling the health impacts of climate systems requires a multidisciplinary approach. The research for this thesis benefited from contact with experts from the fields of geography, climatology, meteorology, and entomology, as well as with public health practitioners.
On the influence of climate and environment on disease transmission

There is sufficient evidence to be confident that medium-term climate indicators (such as the SOI and sea surface temperature) strongly influence the broad-scale rainfall and tidal patterns in parts of Australia. However, their value as predictors of RRv disease epidemics at the regional level is still limited, because of the variable magnitude, spatial characteristics, and duration of ENSO events. More research is needed into the influence of weakly positive and negative, and neutral, values on rainfall distribution in Australia (and hence mosquito breeding). ENSO forecast certainty would also be increased if the length and strength of phase-locking were better understood.

There is much scope for exploring the utility of other environmental factors, both to refine the classification of suitable RRv bioclimatic regions for analysis, and for use in predictive modelling for RRv disease. For example, the NDVI has proven a useful predictor of epidemics of other mosquito-borne diseases (e.g. Rift Valley Fever and malaria). Remotely sensed weather data may be a more useful exposure measurement in less-populated areas of Australia, where the low density of meteorological stations can affect the accuracy of interpolated weather data.

The impact of future climate change on the timing and distribution of RRv disease epidemics is still largely unknown. However, the effect that reduced irrigation in the Murray Valley has had on RRv species type and density in that area (Dhileepan 1996) provides advance notice that projected climate changes (i.e., altered seasonality and variability of rainfall) will affect RRv disease transmission. The research difficulties in this regard are immense. The findings of this thesis, supported by previous observations of virus and mosquito experts in Australia (Lindsay and Mackenzie 1997, Russell 1998b, Mackenzie et al. 2000), suggest that:

- Climate change impacts cannot be assessed for the country as a whole, but need to be analysed within ecologically distinct regions (such as were used in this thesis).
- Decreases in rainfall are an expected feature of climatic change in many parts of Australia. These conditions may have beneficial effects by reducing RRv transmission. At the same time, rainfall is projected to become more extreme, which may increase the occurrence of large outbreaks of disease in some regions.
- Advances in climate change projections – particularly regarding rainfall variability, and changes in the intensity and cycle of ENSO – will assist research.
Further research into the response of the principal RRv vectors to weather conditions (in particular temperature, but also humidity and rainfall) would help improve the assessment of future climate change.

From the point of view of preparing for future change in the incidence and distribution of RRv disease, it would be useful to know which populations will be most vulnerable (Kovats et al. 2001). This could involve the use of existing mosquito and disease surveillance data to detect changes in the patterns of disease and vectors over coming decades.

On the development of mosquito-borne disease early warning systems

Harley and others (2001) have proposed some partial and conservative estimates of the costs of RRv disease, but there is a pressing need to conduct a comprehensive assessment of these costs. Until this occurs, it is not possible to accurately assess the benefits that early warning systems could provide to reduce the level of risk of RRv disease in different regions of Australia. This exercise should include an apportioning of the costs borne by Local Governments, State Health Departments, and communities, as well as the individual and their family. In this manner, the utility of different interventions could be evaluated.

Main conclusions

The main conclusions arising from the original research conducted for this thesis are:

- The analysis of RRv disease for Australia showed that the temporal pattern of disease, the endemic or epidemic nature of infections, and the average annual incidence rate varied substantially across the country, and were strongly determined by long-term average climate. Many patterns of RRv disease (such as inter-annual fluctuations and intra-climate zone variations) cannot be adequately explained at the broad-scale. The ecology of RRv is too complex (in terms of the variety of vertebrate and vector species involved in transmission in different regions) and variable (in terms of the patterns of disease across Australia) to develop accurate models of RRv disease within national, State or broad-scale climate boundaries. The novelty of these findings lie with the length of the RRv disease time series used and the extent of geographic area over which the analyses were conducted.
• Long-term average rainfall, temperature, and vegetation zones can be used to derive smaller ecologically appropriate regions (based on aggregates of population units), within which accurate prediction of RRv disease epidemics can occur.

• The most important contribution of this thesis is the demonstration that climate and environment data can be successfully used to predict epidemics of RRv disease at the regional level in Australia. The sensitivity of the predictive models was reasonable to excellent (62-96%) in the Murray study area, and reasonable in the Southwest (64%). For the southern part of the Southwest, and for both of the Murray regions, this approach substantially improves predictive certainty over what is currently available (i.e., RRv disease surveillance).

• This thesis also demonstrates that mosquito surveillance data, in addition to climate and environment data, can dramatically improve the sensitivity of RRv disease epidemic prediction (from 64% to 90% in the Southwest).

• Importantly, meteorological variables can provide advance warning of epidemics in a timely fashion, so that public health response has an opportunity to reduce the number of cases. The early warning predictive models give at least two months advance notice of epidemics over currently available surveillance techniques.

• This thesis argues that these predictive models could be used as the basis of early warning systems for RRv disease. The advance warning and additional certainty provided by the predictive models could be used to improve the efficiency and effectiveness of mosquito control, and the targeting of public notification campaigns. The cost of conducting surveillance of RRv and its mosquito vectors is substantial, and is only routinely conducted in very few regions of Australia. By signalling high risk regions for data collection, the predictive models could form the basis of a cost-effective allocation of limited resources.

• The methods developed in the two case study areas suggest that climate and environment data could be used in other regions of the country, and for the prediction of other mosquito-borne diseases.

The research questions that shaped this thesis grew out of the environmental health issues that dominated the 1990s. At that time, massive environmental change at global,
national and regional levels was being systematically documented. The Earth Summit in Rio de Janeiro in 1992, and the Second Assessment Report of the IPCC (1995), presented evidence of the unprecedented rate of ecological disruption and climate change and the discernible human influence on these processes. During this period, the diverse pathways by which ecosystem degradation may affect human health were also elucidated (McMichael 1993, IPCC 1995), and evidence of the initial impacts accumulated (National Health and Medical Research Council 1990, McMichael et al. 1996). It was apparent that reducing the total level of greenhouse gas emissions and sustaining the environment must become primary preventive health strategies.

Over the period of this thesis, the Australian Government has taken a trenchant oppositional view to the majority of the global community on emissions targeting and adoption of the Kyoto Protocol (Guest 1997). The social and economic issues that affect the global environment do not appear to have abated, and the urgency for both mitigation and adaptation to environmental change has intensified (IPCC 2001c). The health impacts of future global change will be strongly determined by the effectiveness of a community’s public health system, and the adequacy of its adaptive response. It is a basic function of environmental health practitioners to anticipate, understand, predict, prevent, and monitor environmental health threats (Taylor and Guest 2001). In this context, the main finding of this thesis — of the utility of climate and environment data for early warning of mosquito-borne infections — remains especially relevant for public health response.
Appendix A

AWARD, PUBLICATIONS, AND CONFERENCE PRESENTATIONS

A.1 Award

In conjunction with co-authors, I was awarded the 2002 Rothman Epidemiology Prize. This prize is awarded for the best paper published in the journal *Epidemiology* in 2002. Criteria were importance, originality, clarity of thought, and excellence in writing. Prize of US$3000. The title of the paper was: “Predicting Ross River virus epidemics from regional weather data”.

A.2 Publications

The following papers or chapters were published as a result of the thesis research:


Related publications that build on the work of the thesis include:


### A.3 Relevant presentations at conferences

**International**


**Australia**


References

References


_____. 2002b. *Average weekly earnings: states and Australia (catalogue number 6302.0)*. Canberra, Australia: Australian Bureau of Statistics.


References


Climate Impacts Group. 1996. Climate change scenarios for the Australian region. Melbourne: Division of Atmospheric Research, CSIRO.


References


References


References


Referencias


Gubler DJ. Division of Vector-borne Infectious Diseases, Centers for Disease Control. Fort Collins, Colorado: USA. Personal communication, August 2001.


Hueston L. Head, Arbovirus and Emerging Diseases Unit, Institute of Clinical Pathology and Medical Research, Westmead Hospital. Sydney: Australia. Personal communication, August 2002.


Lindsay M. Medical Entomologist, Mosquito Borne Disease Control Section, Department of Health, Western Australia. Perth: Australia. Personal communication, April 2002.


References


References


344
References


Moodie K. SILO, Queensland Department of Natural Resources. Sydney: Australia. Personal communication, September 1999.


Mott JJ and McComb AJ. 1975. The role of photoperiod and temperature in controlling the phenology of three annual species from an arid region of Western Australia. *Journal of Ecology*; 60:293-304.


References


____. 2003b. *Climate details*. Accessed from: 


Queensland Department of Natural Resources. 2000. *The Data Drill*. Brisbane: Queensland Centre for Climate Applications, Department of Natural Resources.


References


References

Research Organisation, and the Queensland Institute of Medical Research; Vol.4, p.171-174.


References


References


References


