

Identifying hotspots of type 2 diabetes risk using general practice data and geospatial analysis: an approach to inform policy and practice

Nasser Bagheri^{A,G}, Paul Konings^B, Kinley Wangdi^C, Anne Parkinson^B,
 Soumya Mazumdar^D, Elizabeth Sturgiss^E, Aparna Lal^F, Kirsty Douglas^E and
 Nicholas Glasgow^B

^ACentre for Mental Health Research, Research School of Population Health, Australian National University,
 63 Eggleston Road, Acton 2601, Australia.

^BDepartment of Health Services Research and Policy, Research School of Population Health,
 Australian National University, 62 Eggleston Road, Acton, ACT 2601, Australia.

^CDepartment of Global Health, Research School of Population Health, Australian National University,
 62 Eggleston Road, Acton, ACT 2601, Australia.

^DHealthy People and Place Unit, Population Health, Liverpool Hospital, South West Sydney Local Health
 District, New South Wales Health, 52 Scrivener Street, Warwick Farm, NSW 2170, Australia.

^EDepartment of General Practice, Monash University, 270 Ferntree Gully Road, Notting Hill,
 Vic. 3168, Australia.

^FNational Centre for Epidemiology and Population Health, Research School of Population Health,
 Australian National University, 62 Eggleston Road, Acton, ACT 2601, Australia.

^GCorresponding author. Email: nasser.bagheri@anu.edu.au

AQ1

Abstract. The prevalence of type 2 diabetes (T2D) is increasing worldwide and there is a need to identify communities with a high-risk profile and to develop appropriate primary care interventions. This study aimed to predict future T2D risk and identify community-level geographic variations using general practices data. The Australian T2D risk assessment (AUSDRISK) tool was used to calculate the individual T2D risk scores using 55 693 clinical records from 16 general practices in west Adelaide, South Australia, Australia. Spatial clusters and potential 'hotspots' of T2D risk were examined using Local Moran's I and the Getis-Ord G_i^* techniques. Further, the correlation between T2D risk and the socioeconomic status of communities were mapped. Individual risk scores were categorised into three groups: low risk (34.0% of participants), moderate risk (35.2% of participants) and high risk (30.8% of participants). Spatial analysis showed heterogeneity in T2D risk across communities, with significant clusters in the central part of the study area. These study results suggest that routinely collected data from general practices offer a rich source of data that may be a useful and efficient approach for identifying T2D hotspots across communities. Mapping aggregated T2D risk offers a novel approach to identifying areas of unmet need.

Additional keywords: geographical variation, primary health care, spatial clusters, T2D risk.

Received 28 February 2019, accepted 23 August 2019, published online xx xxxxx xxxx

Introduction

Type 2 diabetes (T2D) is now ranked in the top four high-priority, non-communicable diseases (NCDs) by the [World Health Organization \(2016\)](#). In Australia, T2D sits within the top 20 causes of total disease burden, accounts for more than A\$1.5 billion of healthcare expenditure annually and has been recognised as a national health priority area since 1997 ([Australian Institute of Health and Welfare 2016a, 2016b, 2016c](#)). There is an increasing need for targeted community-based interventions that can prevent T2D. Early identification of people who have a higher risk of developing T2D and

geographical clusters (hotspots) of T2D risk will help policy-makers to tailor and target preventative interventions to communities where T2D risk is greatest.

Numerous T2D risk calculation tools for determining the probability of developing T2D over a 5- to 10-year period have been developed worldwide, with seven tools rated highly for use in routine clinical practice. They include AUSDRISK (Australia), QDScore and Cambridge Risk Score (UK), FINDRISC (Finland), ARIC (Germany) and the Framingham Offspring Study and San Antonio Risk Score (USA) ([Noble et al. 2011; Abbasi et al. 2012; Kilkenny et al. 2014](#)). The

What is known about the topic?

- There is strong evidence that lifestyle, nutrition and pharmacological interventions can reduce the risk of developing type 2 diabetes (T2D) in individuals.

What does this paper add?

- Advanced geospatial analysis of T2D risk identified areas of hotspots and cold spots in the study communities, and this can be used to inform policy planning and develop tailored interventions.

AUSDRISK tool developed in 2008 has been validated and recommended for use in Australian settings (Grant *et al.* 2006; Cugati *et al.* 2007; Australian Government Department of Health 2008; Tonkin *et al.* 2009; Pasco *et al.* 2010; Noble *et al.* 2011). It estimates the risk of an individual developing T2D using 10 risk factors including age, gender, ethnicity, family history, blood glucose, blood pressure, smoking, physical activity, diet and waist circumference (Australian Government Department of Health 2008).

The value of spatial analysis in public health is well recognised and is a useful approach to highlight clusters or 'hotspots' of T2D risk across communities (Curtis and Lee 2010). For example, geospatial mapping of a national US survey identified a high prevalence 'T2D belt' across southern USA states. A recent systematic review of spatial modelling of T2D outcomes found that, although several spatial studies had been conducted in the USA, UK and Europe, there is a lack of similar studies in other parts of the world (Baker *et al.* 2015). Of the included 10 studies, only a single study used general practice data (Noble *et al.* 2012). Few studies have been undertaken in this emerging area, combining general practice data and spatial analysis at the small area level. A study by Noble *et al.* (2011) examined the feasibility of mapping T2D risk at the community level in a London (UK) district. In Australia, several small studies have used general practice data to identify and map areas of undiagnosed T2D (Bagheri *et al.* 2014), diagnosed T2D (Jiwa *et al.* 2015) and cardiovascular disease (CVD) risk (Bagheri *et al.* 2015) and, in the USA, a larger study by Gabert *et al.* (2016) examined diagnosed T2D. However, this unique and rich source of routinely collected data in general practice is underutilised for the purposes of T2D risk prediction, geospatial analyses and community-based interventions.

The aims of this study were to: (1) calculate the probability of developing T2D risk in individuals using the AUSDRISK (Baker IDI Heart and Diabetes Institute, Melbourne, Vic., Australia) tool and routinely collected data from general practices; (2) assess spatial variations of T2D risk; and (3) identify hotspots at the small area level and their association with the socioeconomic background of communities in Australia.

Methods*Settings and study population*

We used a de-identified cross-sectional cohort of data from 16 general practices located in western Adelaide (South Australia)

for the period 1 January 2012 to 31 December 2014. We included patients meeting the following criteria: aged ≥ 18 years ($n = 55\,693$) and who were active patients (defined as having had three visits to a GP within the last 2 years). There was a total of 133 363 patient records. We excluded patients diagnosed with type 1 or T2D ($n = 4288$) (see Figure S1 available as Supplementary Material to this paper). The data were extracted from general practices' records using the PEN CAT clinical audit tool (PenCS, Sydney, NSW, Australia). PEN CAT searches coded reasons for presentation, smoking and alcohol history, medications and pathology results.

The T2D risk estimation was carried out using the AUSDRISK tool. Eight of the 10 risk factors were considered; age, sex, ethnicity, blood sugar level, taking medication for hypertension, smoking, waist circumference and regular weekly physical activity. Family history of T2D and fruit and vegetable consumption were not available.

Missing values for certain variables were assigned an imputed or default value following the methodology used by Noble *et al.* (2012) and Mathur *et al.* (2012) to estimate the risk of developing T2D in East London using general practice data and the QDScore (Hippisley-Cox *et al.* 2009).

Missing data on smoking was imputed to be non-smoking data. If ethnicity was not stated, the algorithm defaulted to non-Indigenous. Other predictor variables were only recorded if positive, such as taking antihypertensive medications (Mathur *et al.* 2012; Noble *et al.* 2012). Waist circumference was imputed using age, sex, systolic BP and status of having diabetes or not, which was correlated with waist circumference from a pre-defined set of potentially contributing variables following the approach taken by the Australian Bureau of Statistics (ABS) National Health Survey (2017–18) (Australian Bureau of Statistics 2018). Multiple imputations occurred before excluding cases with diagnosed diabetes from the analysis. We predicted type 2 diabetes risk for those who had not been already diagnosed with T2D. There were no missing values for age, sex and diabetes status, and only 12% were missing for systolic blood pressure (SBP) values. However, the majority of missing values for SBP were seen in younger age participants (<40 years), who usually have a lower risk of T2D compared with people aged >45 years.

The risk of developing T2D within 5 years was classified as: 5 or less: low risk; 6–11: intermediate risk; and 12 or more: high risk (Australian Government Department of Health 2008). Additionally, a sensitivity analysis was performed by generating the diabetic risk score without age to determine clustering trends.

General practice records were geocoded at statistical area level one (SA1) using methods described by Mazumdar *et al.* (2014). SA1 is the smallest area of output used for the ABS census of population and housing, representing an average population of ~400 persons. SA1s with a population of less than six patients were excluded from the final analysis to ensure patient confidentiality.

Data were extracted for Socio-Economic Indexes for Areas (SEIFA) (Public Health Information Development Unit 2016) (Fig. S2), developed by the ABS to rank areas in Australia according to relative socioeconomic advantage and disadvantage (Australian Bureau of Statistics 2013). The data were then linked to general practice data to examine the relationship between T2D risk and the socioeconomic background of communities.

Geospatial analysis

Individual T2D risk scores were summarised to the SA1 level to examine spatial variation of mean T2D risk across communities in the study area. Spatial hotspots at the SA1 level with high or low T2D risk were examined using the Getis-Ord Gi* technique (Getis and Ord 1992; Ord and Getis 1995). This tool compares the local sum of mean T2D risk scores (the sum of the mean T2D risk of the targeted SA1 area and its neighbouring SA1s) to the sum of mean T2D risk scores of all SA1s within the study area. A statistically significant large, positive Z-score signifies a local high-T2D risk cluster (hotspot). Hotspots are detected when SA1s with high T2D risk are surrounded by SA1s with high T2D risk and when the observed local sum of T2D risk is higher than the expected local sum, and the difference is too large to be the result of chance alone. Similarly, a statistically significant large, negative Z-score signifies a local low-risk cluster (cold spot), where SA1s with low T2D risk are surrounded by other SA1s with low risk of T2D (Getis and Ord 1992; Anselin 1995; Ord and Getis 1995). Clusters of statistically significant hotspots and cold spots were visualised in the western Adelaide area to highlight communities with a high or low risk respectively.

We also used the Local Moran's I statistic to identify local spatial clusters of SA1-level T2D risk (Anselin 1995). The statistic identifies hotspots (high-high), cold spots (low-low) and spatial outliers (high-low and low-high). A positive Local Moran's I value indicates that the target SA1 is surrounded by SA1s with similar risk values (high-high: SA1 with a high risk surrounded by SA1s with high T2D risk; low-low: SA1 with a low risk surrounded by SA1s with low risk). A negative Local Moran's I value indicates that the target SA1 is surrounded by SA1s with dissimilar risk (high-low: SA1 with a high risk surrounded by SA1s with low risk; low-high: SA1 with a low risk surrounded by SA1s with high risk) (Anselin 1995). The designation of SA1s to these four classes was determined using a statistical test that performs random comparisons with Local Moran's I values among the target SA1s and neighbouring SA1s with Local Moran's I values of all SA1s within the study area, and compares the observed Local Moran's I value to the value corresponding to the random permutations (expected Local Moran's I value) (Anselin 1995). If the test is significant ($P \leq 0.05$), the observed Local Moran's I value is significantly larger (or smaller in the case of a negative relationship) than the expected Local Moran's I value. If the test is not significant, the SA1 remains in a neutral class (no spatial dependence) (Anselin 1995). Statistically significant high-high, low-low and outlier local clusters were visualised using a map with SA1 boundaries. Sensitivity analysis of T2D risk and its clustering was undertaken by calculating T2D risk with and without including age in the prediction models. In addition, a Global Moran's I was computed, which provides an indication of whether a hotspot exists in the study area. Unlike the Local Moran's I, this statistic does not pinpoint the location of the hotspot. A summary of activities can be seen in Figure 1.

We used Stata (version 14.2; StataCorp LP, College Station, TX, USA) to calculate T2D risk scores and conduct statistical analyses, and ArcGIS (version 10.2; ESRI, Redlands, CA, USA) to conduct spatial analyses. Ethical approval was granted by

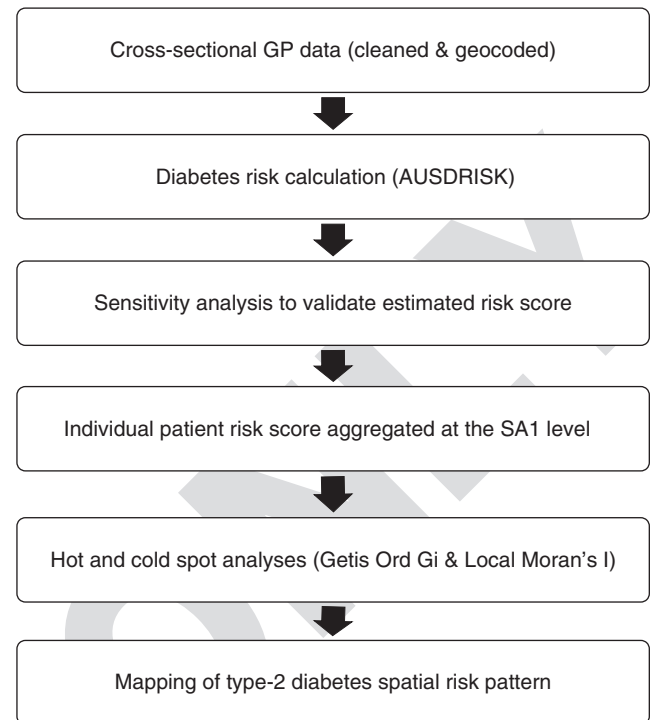


Fig. 1. Summary of activities workflow diagram.

the Australian National University Human Ethics Committee (Protocol 2014/174).

Results

Study population characteristics

The risk of developing type 2 T2D within 5 years at the individual level (Table 1) showed that 34.0% (18,943) were at low risk (≤ 5 points); 35.2% (19,574) were at intermediate risk (6–11 points); and 30.8% (17,176) were at high risk (≥ 12 points). A sensitivity analysis was also undertaken by calculating the T2D risk score using non-imputed values and imputed values separately in the risk prediction model. Table 2 shows that the results are comparable with a sufficiently small impact.

T2D risk score and geospatial analysis

A total of 467 SA1s with a population of 54,482 were included in the geospatial analysis. The number of active patients varied from 10 to 481, with a mean of 116.7 across the SA1s. The T2D risk scores were heterogeneous across the SA1s, with scores ranging from 5.37 to 12.93 (standard deviation of 1.1; Fig. 2).

The Global Moran's I showed that there were significant spatial autocorrelations (z score = 10.916604; $P < 0.0001$), indicating the presence of one or more hotspots in the study area (Fig. S3). Our maps demonstrated that T2D risk was higher in the eastern and central parts of the study area. Using the Getis-Ord Gi* technique, we found significant hotspots in the central parts, whereas cold spots were located in the northern and south-eastern parts of the study area (Fig. 3). These findings were supported with the analysis of clusters and outliers using Anselin's Local Moran's I. There were 55 SA1s with clusters

AQ6

Table 1. Characteristics categorised by risk of type 2 diabetes (T2D) within 5 years
Data are given as n (%)

| Characteristics | Type 2 diabetes risk (%) with 5 years | | |
|---|---------------------------------------|----------------------------------|-------------------------|
| | Low (score ≤5) (n = 18 943) | Intermediate (6–11) (n = 19 574) | High (≥12) (n = 17 176) |
| Age groups (years) | | | |
| 18–34 | 10 768 (56.8) | 3894 (19.9) | 161 (0.9) |
| 35–44 | 5421 (28.6) | 3351 (17.1) | 555 (3.2) |
| 45–54 | 2754 (14.5) | 4544 (23.2) | 2761 (16.1) |
| 55–64 | 0 | 4072 (20.1) | 4379 (25.5) |
| ≥65 | 0 | 3716 (19.0) | 9317 (54.3) |
| Gender | | | |
| Male | 5547 (29.3) | 8086 (41.3) | 942 (55.0) |
| Female | 13 396 (70.7) | 11 491 (58.7) | 7731 (45.0) |
| Ethnicity | | | |
| Aboriginal, Torres Strait Islander (ATSI) | 8 (0.04) | 168 (0.9) | 370 (2.2) |
| Not ATSI | 18 935 (99.9) | 19 409 (99.1) | 16 803 (97.8) |
| High blood glucose | | | |
| No | 18 943 (100.0) | 11 679 (59.7) | 2118 (12.3) |
| Yes | 0 (0.0) | 7895 (40.3) | 15 058 (87.7) |
| Currently taking medication for high blood pressure | | | |
| No | 18 845 (99.5) | 16 778 (85.7) | 10 649 (62.0) |
| Yes | 98 (0.5) | 2796 (14.3) | 6527 (38.0) |
| Currently smoking cigarettes or any other tobacco products on a daily basis | | | |
| Smoking | 17 272 (91.2) | 17 106 (87.4) | 15 011 (87.4) |
| Never smoked | 1671 (8.8) | 2468 (12.6) | 2165 (12.6) |
| Physical activity | | | |
| No | 2 (100.0) | 8 (66.7) | 39 (56.5) |
| Yes | 0 | 4 (33.3) | 30 (43.5) |

Table 2. Type 2 diabetes (T2D) risk score of imputed and non-imputed

| | Low | Intermediate | High |
|-----------------|---------------|---------------|---------------|
| Imputed | | | |
| N (%) | 18 943 (34.0) | 19 577 (35.2) | 17 173 (30.8) |
| Mean (min, max) | 2.2 (0, 5) | 8.4 (6, 11) | 15.7 (12, 28) |
| Non-Imputed | | | |
| N (%) | 12 899 (23.1) | 26 136 (46.7) | 16 903 (30.2) |
| Mean (min, max) | 1.0 (2, 5) | 8.3 (6, 11) | 15.1 (12, 30) |

of high–high and 53 with low–low clusters, nine were low–high outliers and a further six were high–low outliers (Fig. 4). Validation of T2D hotspots without age showed a similar trend (Fig. S4–S6).

5 **Predicted T2D risk and SEIFA**

The T2D risk score ranged from 4.1 to 10.9 in relation to SEIFA quartiles (Table 3). The Pearson correlation analysis showed that there was a weak significant negative association between T2D risk and the socioeconomic background of communities in the study ($r = -0.0879$, $P = 0.0476$) (see Table S1 available as Supplementary Material to this paper). Moreover, the linear regression model supported the reverse relationship between T2D risk and SEIFA; with an increased socioeconomic level, the T2D risk decreases (-0.26 with 95% CI -0.5118821 to -0.0250103).

Using general practice clinical records, we predicted the probability of developing T2D risk and identified communities with T2D risk clusters. When these clusters (hotspots) were examined by socioeconomic status at the SA1 level, we found a significant inverse association.

Discussion

Whereas in Australia there are extant T2D risk estimates at small area geographies, these are produced at relatively coarse geographies using ecological models and census data (Public Health Information Development Unit 2016). Our methods used individual level data and produced risk estimates at very high geographic resolutions, greatly increasing the accuracy and usability of our estimates.

By applying the AUSDRISK tool to general practice data from western Adelaide, this study estimated the predicted 5-year risk of developing T2D by selected characteristics and by where people live. We found that three in 10 people in our sample were at high risk of developing T2D within 5 years. This finding was in line with the Centers for Disease Control and Prevention report, which estimated 35% of the USA population aged ≥20 years were at high risk of developing T2D based on fasting glucose or hemoglobin A1c levels (Centers for Disease Control and Prevention 2011).

The results indicate that the predicted 5-year risk significantly increases with age, especially for those aged ≥64 years; a similar finding to that of Hippisley-Cox et al. (2009), who also used the QDScore tool in England and Wales to predict the

5

10

15

20

25

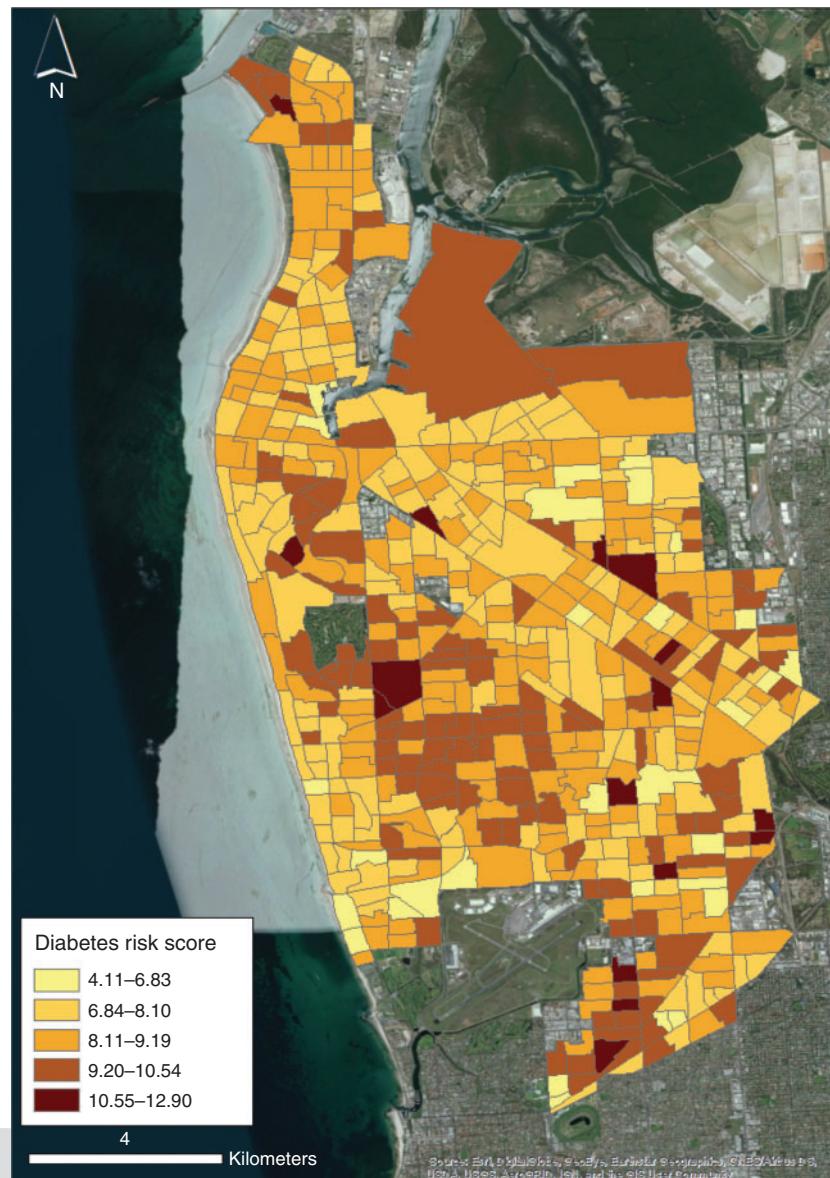


Fig. 2. Type 2 diabetes risk map at statistical area level 1 (SA1) level.

10-year risk of T2D. As with most chronic conditions, the prevalence is higher in older adults, with the majority of those with T2D in developed countries aged ≥ 65 years (Gruneir *et al.* 2016). Australia has a growing ageing population and this causes a significant burden on the healthcare system. Other contributing factors include smoking, lack of physical exercise and high blood glucose levels. Another indicator may be taking medication for hypertension, which was more commonly reported (59%) in men at very high risk of developing T2D (AUSDRISK score ≥ 20) in a study of Australian men conducted by Aguiar *et al.* (2015). Males were also found to be at higher risk, which is in line with the literature about male sex as a risk factor for the development of T2D. The findings also support predictions that one-third of young Australians will develop T2D in their lifetime (Australian Institute of Health and Welfare

2016a, 2016c). As it is now well established that T2D is a major risk factor for CVD and stroke, addressing T2D would therefore have a flow-on effect for CVD and stroke.

The aim of this study was to predict future risk of T2D using general practice clinical data and investigate potential spatial variation at the community level. Our findings showed a high level of spatial variation in T2D risk across communities. National-level policymakers are interested in developing community-based interventions to manage chronic disease, hence, it is of the utmost importance that we are able to identify communities at the highest risk of T2D so that appropriate intervention programs may be designed based on the community T2D risk profile.

Our approach offers a new means of identifying those communities most at risk of developing T2D and can guide

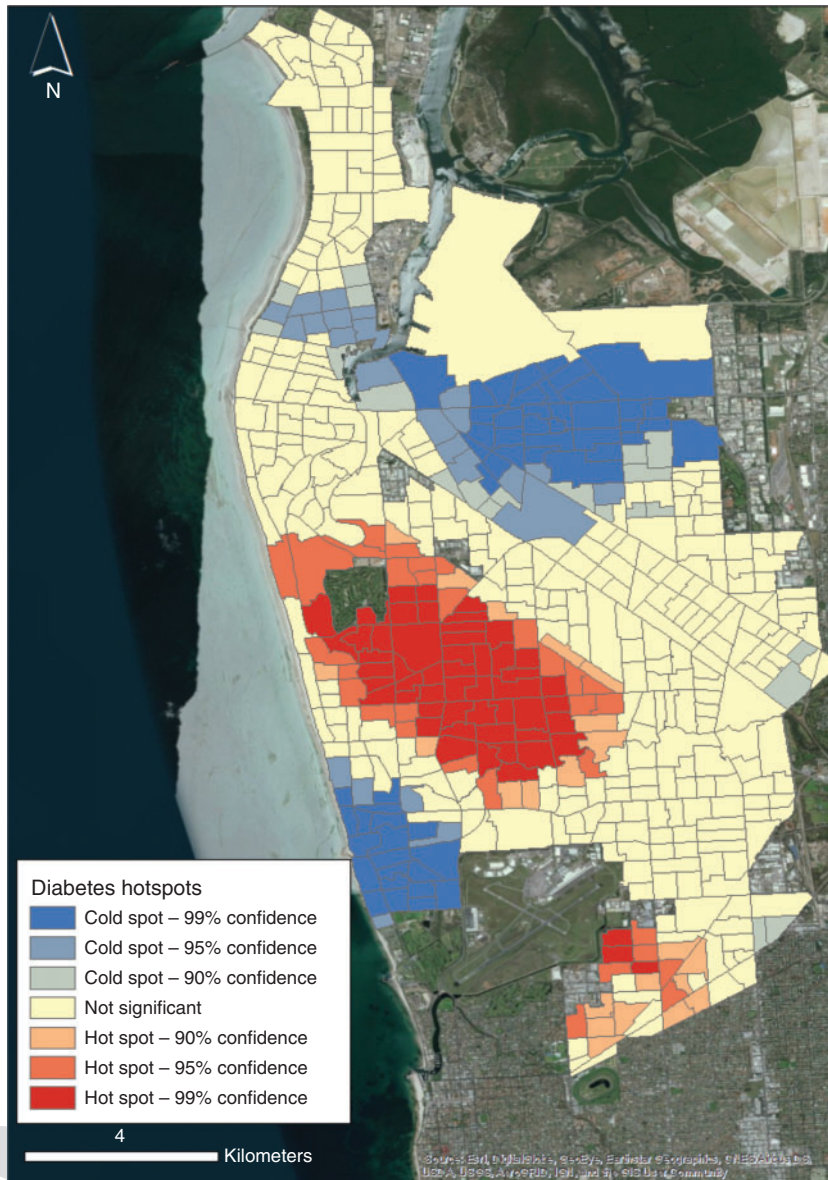


Fig. 3. Type 2 diabetes risk hotspot map at statistical area level 1 (SA1) level.

the design of prevention strategies for use in patient-centred, community-based T2D risk factor modification. The large size of our study population and the detailed general practice clinical data of patients mapped in a relatively confined geographic area make our study unique to date. The methods proposed in this research can be easily applied to other datasets in different jurisdictions with appropriate data infrastructure. Using spatial analysis, we were able to identify a heterogeneity pattern and significant hotspots in the central parts of Adelaide, whereas overall T2D risk was found to be higher in the eastern and central parts compared with northern and south-eastern parts.

We expected to find hotspots in the north-east neighbourhoods, which are low socio-economic status (SES) areas; however, our analysis showed a cold spot in this region. The

reasons for this are two-fold: a very weak negative correlation of SES and diabetes risk can be observed ($r = -0.08$, $P = 0.047$); and people living in this low socioeconomic area were generally younger (47 years in this area compared to 53 years in hotspots areas), had a higher normal range of blood sugar and lower body mass index.

Strengths and limitations

The clinical data on which this study was based were drawn from a defined area within Adelaide. As such, it is possible that our conclusions are only applicable to this area. This new method would benefit from further application to other sites. Not all 10 variables from the AUSDRISK tool were used. Fruit and vegetable consumption were not recorded in general

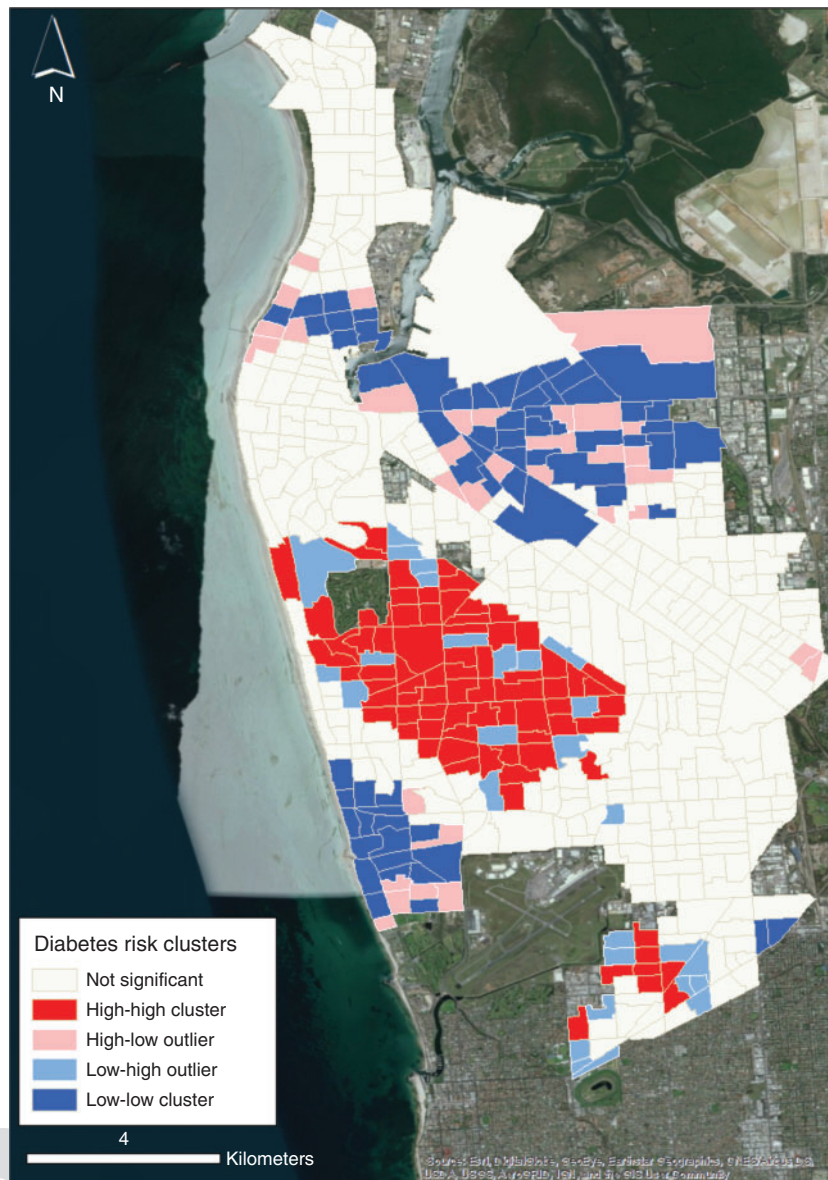


Fig. 4. Type 2 diabetes risk clusters and outlier map (Anselin Local Moran's I) at statistical area level 1 (SA1) level.

Table 3. Socio-Economic Indexes for Areas (SEIFA) with the predicted risk of type 2 diabetes (D)

| SEIFA tertile | Type 2 diabetes risk score (%) | | | |
|---------------|--------------------------------|------|------|-------------------------|
| | Mean | Min. | Max. | 95% confidence interval |
| Low | 8.6 | 4.1 | 12.9 | (8.4, 8.8) |
| Moderate | 8.6 | 5.2 | 11.5 | (8.4, 8.7) |
| High | 8.3 | 5.0 | 10.9 | (8.2, 8.5) |

practice data, thus we set the weight of these two variables to zero by default in the risk assessment tool. We imputed waist circumference using a multiple imputation technique.

Further, we set the weight of physical activity for those patients who did not state their physical activity to zero by default. This may slightly underestimate T2D risk for those who do not have a record of physical activity status in the general practice dataset.

The AUSDRISK tool is validated for use in Australia. Variables are collected as part of routine clinical practice and recorded within an individual's patient record, which avoids bias associated with self-reporting. Using existing general practice data routinely collected for clinical purposes is simple and efficient. The study drills down to a fine-grained geographic level, and incorporating hotspots analysis and socioeconomic status can further assist with geographic targeting of prevention measures to the communities at greater risk.

Conclusion

We have shown that aggregated T2D risk offers a novel approach for identifying high-risk neighbourhoods that are missed by the larger regions identified in health examination surveys. Our results suggest that routinely collected general practice data may provide a rich resource for identifying T2D hotspots at the community level.

Using spatial analysis techniques in tandem with the AUSDRISK risk calculation tool, this study estimates risk prediction at the small area level (community) and provides essential information in the form of a visual snapshot of where to target preventive interventions. The areas with the highest predicted risk need more interventions to effectively reduce outcome disparities.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

This article received funding support from the Australian Research Council (DE14 0101570). We thank the 16 west Adelaide (Australia) general practices that provided data for this study.

References

- Abbasi A, Peelen LM, Corpeleijn E, van der Schouw YT, Stolk RP, Spijkerman AMW, van der A DL, Moons KGM, Navis G, Bakker SJL, Beulens JWJ (2012) Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ (Clinical Research Ed.)* **345**, e5900. doi:10.1136/bmj.e5900
- Aguiar EJ, Morgan PJ, Collins CE, Plotnikoff RC, Callister R (2015) Characteristics of men classified at high-risk for type 2 diabetes mellitus using the AUSDRISK screening tool. *Diabetes Research and Clinical Practice* **108**, 45–54. doi:10.1016/j.diabres.2015.01.017
- Anselin L (1995) Local indicators of spatial association—LISA. *Geographical Analysis* **27**, 93–115. doi:10.1111/j.1538-4632.1995.tb00338.x
- Australian Bureau of Statistics (2013) Socio-Economic Indexes for Areas (SEIFA). (ABS, Canberra, ACT, Australia) Available at <https://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa> [Verified 6 November 2019]
- Australian Bureau of Statistics (2018) 4364.0.55.001 – National Health Survey: First Results, 2017–18. (ABS, Canberra, ACT, Australia) Available at <https://www.abs.gov.au/ausstats/abs@.nsf/mf/4364.0.55.001> [Verified 6 November 2019]
- Australian Government Department of Health (2008) Australian type 2 diabetes risk assessment tool (AUSDRISK). (Department of Health: Canberra, ACT, Australia) Available at <http://www.health.gov.au/preventionoftype2diabetes> [Verified 22 August 2019]
- Australian Institute of Health and Welfare (2016a) National health priority areas. (AIHW: Canberra, ACT, Australia) Available at <http://www.aihw.gov.au/national-health-priority-areas/> [Verified 8 August 2009]
- Australian Institute of Health and Welfare (2016b) Australia's health 2016. Australia's health series number 15. Catalogue number AUS 199. (AIHW: Canberra, ACT, Australia) Available at <http://www.aihw.gov.au/publication-detail/?id=6012955544> [Verified 15 September 2009]
- Australian Institute of Health and Welfare (2016c) Australian Burden of Disease Study: impact and causes of illness and death in Australia 2011. Australian Burden of Disease Study series number 3. Catalogue number BOD 4. (AIHW: Canberra, ACT, Australia) Available at <http://www.aihw.gov.au/publication-detail/?id=60129555173> [Verified 17 August 2009]
- Bagheri N, McRae I, Konings P, Butler D, Douglas K, Del Fante P, Adams R (2014) Undiagnosed diabetes from cross-sectional GP practice data: an


- approach to identify communities with high likelihood of undiagnosed diabetes. *BMJ Open* **4**, e005305. doi:10.1136/bmjopen-2014-005305
- Bagheri N, Gilmour B, McRae I, Konings P, Dawda P, Del Fante P, van Weel C (2015) Community cardiovascular disease risk from cross-sectional general practice clinical data: a spatial analysis. *Preventing Chronic Disease* **12**, 140379. doi:10.5888/pcd12.140379
- Baker J, White N, Mengersen K (2015) Spatial modelling of type II diabetes outcomes: a systematic review of approaches used. *Royal Society Open Science* **2**, 140460. doi:10.1098/rsos.140460
- Centers for Disease Control and Prevention (CDC) (2011) National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011. (CDC: Atlanta, GA, USA)
- Cugati S, Wang JJ, Rochtchina E, Mitchell P (2007) Ten-year incidence of diabetes in older Australians: the Blue Mountains Eye Study. *The Medical Journal of Australia* **186**, 131–135. doi:10.5694/j.1326-5377.2007.tb00836.x
- Curtis AJ, Lee WA (2010) Spatial patterns of diabetes related health problems for vulnerable populations in Los Angeles. *International Journal of Health Geographics* **9**, 43. doi:10.1186/1476-072X-9-43
- Gabert R, Thomson B, Gakidou E, Roth G (2016) Identifying high-risk neighborhoods using electronic medical records: a population-based approach for targeting diabetes prevention and treatment interventions. *PLoS One* **11**, e0159227. doi:10.1371/journal.pone.0159227
- Getis A, Ord JK (1992) The analysis of spatial association by use of distance statistics. *Geographical Analysis* **24**, 189–206. doi:10.1111/j.1538-4632.1992.tb00261.x
- Grant JF, Chittleborough CR, Taylor AW, Dal Grande E, Wilson DH, Phillips PJ, Adams RJ, Cheek J, Price K, Gill T, Ruffin RE (2006) The North West Adelaide Health Study: detailed methods and baseline segmentation of a cohort for selected chronic diseases. *Epidemiologic Perspectives & Innovations* **3**, 4. doi:10.1186/1742-5573-3-4
- Gruneir A, Markle-Reid M, Fisher K, Reimer H, Ma X, Ploeg J (2016) Comorbidity burden and health services use in community-living older adults with diabetes mellitus: a retrospective cohort study. *Canadian Journal of Diabetes* **40**, 35–42. doi:10.1016/j.cjcd.2015.09.002
- Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P (2009) Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. *BMJ (Clinical Research Ed.)* **338**, b880. doi:10.1136/bmj.b880
- Jiwa M, Gudes O, Varhol R, Mullan N (2015) Impact of geography on the control of type 2 diabetes mellitus: a review of geocoded clinical data from general practice. *BMJ Open* **5**, e009504. doi:10.1136/bmjopen-2015-009504
- Kilkenny MF, Johnson R, Andrew NE, Purvis T, Hicks A, Colagiuri S, Cadilhac DA (2014) Comparison of two methods for assessing diabetes risk in a pharmacy setting in Australia. *BMC Public Health* **14**, 1227–1237. doi:10.1186/1471-2458-14-1227
- Mathur R, Noble D, Smith D, Greenhalgh T, Robson J (2012) Quantifying the risk of type 2 diabetes in East London using the QDScore: a cross-sectional analysis. *The British Journal of General Practice* **62**, e663–e670. doi:10.3399/bjgp12X6656793
- Mazumdar S, Konings P, Hewett M, Bagheri N, McRae I, Del Fante P (2014) Protecting the privacy of individual general practice patient electronic records for geospatial epidemiology research. *Australian and New Zealand Journal of Public Health* **38**, 548–552. doi:10.1111/1753-6405.12262
- Noble D, Mathur R, Dent T, Meads C, Greenhalgh T (2011) Risk models and scores for type 2 diabetes: systematic review. *BMJ (Clinical Research Ed.)* **343**, d7163. doi:10.1136/bmj.d7163
- Noble D, Smith D, Mathur R, Robson J, Greenhalgh T (2012) Feasibility study of geospatial mapping of chronic disease risk to inform public health commissioning. *BMJ Open* **2**, e000711. doi:10.1136/bmjopen-2011-000711
- Ord JK, Getis A (1995) Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis* **27**, 286–306. doi:10.1111/j.1538-4632.1995.tb00912.x


- Pasco JA, Kotowicz MA, Henry MJ, Nicholson GC (2010) Evaluating AUSDRISK for predicting incident diabetes in an independent sample of women. *The Medical Journal of Australia* **193**, 374. doi:10.5694/j.1326-5377.2010.tb03955.x
- 5 Public Health Information Development Unit (2016) Social Health Atlases. (Torrens University Australia: Adelaide, SA, Australia) Available at <http://phidu.torrens.edu.au/social-health-atlases> [Verified 26 August 2009]
- Tonkin AM, Boyden AN, Colagiuri S (2009) Maximising the effectiveness and cost-effectiveness of cardiovascular disease prevention in the general population. *The Medical Journal of Australia* **191**, 300–302. doi:10.5694/j.1326-5377.2009.tb02809.x
- World Health Organization (WHO) (2016) Global report on diabetes. (WHO: Geneva, Switzerland) Available at http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf [Verified 1 August 2009]

PROOF ONLY

AUTHOR QUERIES

AQ1: ANU addresses – they vary between 62 and 63 Eggleston Road, is this correct? 

AQ2: City has been added as Sydney for PenCS, correct? 

AQ3: Citation to Hippisley-Cox *et al.* 2009 added here for origin of QDScore – correct? If not, please provide the manufacturer details for QDScore. 

AQ4: ‘Local’ added before Moran’s I for consistency of this term throughout your paper – correct? (‘...and neighbouring SA1s with **Local** Moran’s I values...’) 

AQ5: As above re ‘Local’. (‘expected **Local** Moran’s I value’) 

AQ6: Data definition included in headnote For Table 1 – correct definition? 

PROOF ONLY