Estimating diabetes prevalence in South Auckland: how accurate is a method that combines lists of linked health datasets?

Simon Thornley, Roger Marshall, Gary Jackson, James Smith, Wing-Cheuk Chan, Craig Wright, Dudley Gentles, Rod Jackson

Abstract

Aims To assess the accuracy of a method for estimating adult diabetes prevalence that combines linked, routine health datasets in South Auckland, New Zealand.

Methods We used a simple algorithm that combined records of laboratory testing, drug dispensing and hospital diagnoses applied to National Health Index-linked health data in South Auckland to estimate the prevalence of diabetes in 2007. We investigated the sensitivity of this ‘combined list’ algorithm against a gold standard of individuals with diagnosed diabetes enrolled in a Chronic Care Management programme (CCMP). We also assessed the sensitivity of this algorithm against an estimated diabetes population generated using capture-recapture methods.

Results From the combined-list algorithm, 25,797 (7.2%) South Aucklanders aged 15 years and over had diabetes. During this period, 10,725 patients were enrolled in the CCMP. The combined list algorithm correctly identified (sensitivity) 10,351/10,725 (96.5%) of those enrolled. When we used the capture-recapture estimated diabetes population as an alternative gold standard, 34,418 [9.5%] of South Aucklanders 15 years and over had diabetes, with the sensitivity of the combined list method falling to about 75% (25,797/34,418).

Conclusion Linked health data provide reasonably accurate estimates of diabetes prevalence in a New Zealand population; particularly for cases with longstanding or complicated disease.

With the global rise of obesity and Type 2 diabetes in the last 30 years in New Zealand and elsewhere, timely and accurate estimates of the burden of this disease are required to plan health interventions and assess their influence on disease prevalence. In areas where prevalence of diabetes is thought to be high, reliable estimates are required.

In New Zealand, the highest concentration of diabetes is likely to be in Counties Manukau District Health Board (CMDHB), given the high proportion of Māori and Pacific peoples; and the high level of socioeconomic deprivation in this district health board (DHB) population. This administratively defined region encompasses most of South Auckland, and is one of 21 geographically defined DHBs in New Zealand, with a population of about 464,500.

Accurate, cost-effective and timely measurement of diabetes prevalence in CMDHB would help determine the effect of prevention programmes by allowing time trends to be studied. Such information has been extrapolated from national health status.
surveys that allow regional estimates to be calculated, by assuming nationally aggregated ethnic specific rates of disease apply to a local population. Such methods are, however, limited in a number of ways. For example, health status surveys are potentially biased by high levels of non-response (e.g. the 2006/07 New Zealand Health Survey had a non-response rate of 32% in adults\(^2\)). People who do not respond to such surveys, generally, have poorer health than responders. Moreover, the national survey is only carried out every 4 to 5 years, so that few measurements are available for time trend analyses.

Analysis of health care data may more accurately estimate disease prevalence than survey methods, given that its quality has recently improved substantially. From 2005, the proportion of records linked to the National Health Index (NHI)—a unique national individual identifier—has been greater than 90%. Linked health data now allows us to study population trends in health care in New Zealand by combining records of hospital diagnoses from admissions, drug dispensing and laboratory test events, along with mortality. Linked data of primary care enrolments and outpatients attendances are also available, although the coverage is less comprehensive.

These data can generate population prevalence estimates of diabetes using two alternative methods. The first is an individual “combined list estimate”, or algorithm, that counts the number of people with diabetes-related activities appearing on different databases (diagnosis at hospital discharge, history of laboratory tests, or pharmaceutical dispensing). An alternative approach uses model based ‘capture-recapture’ methods that can estimate the extent of undercount associated with the ‘combined-list’ procedure.\(^3,4\)

The strength of the former (combined list) approach is that it requires fewer assumptions to be made, and simply counts individuals who are recorded with indicators of diabetes related care in health service databases. It is more likely to identify people with more serious conditions that result in more contact with health services.

The alternative (capture-recapture) approach has some important advantages over the combined list approach but also important weaknesses. The strength of capture-recapture methods is their potential to estimate the number of people who are missing from databases, either because of incomplete recording or because they have never been identified. Their main weakness, however, is the requirement to assume that the counts from lists, along with intersections between lists, conform to specific probability distributions (Poisson) specified in statistical (log-linear) models.

Therefore capture-recapture estimates need to be interpreted with considerable caution and we have used this method as a secondary validation tool, rather than as our primary method, for estimating diabetes prevalence.

**Methods**

**Background and study design**—We used routine health data sources to calculate the prevalence of diabetes in residents of CMDHB, identifying the same individual on each database by their NHI number. Every New Zealander has a unique NHI alphanumeric code that is encrypted and used to anonymously link different databases. This identifier is now linked to most routinely collected national health databases. Appearance in at least one of these databases defined a denominator population, from which, cases of diabetes are presumed to occur. Then, cases of diabetes were drawn from a subset of this population.
To define the denominator population, we selected people who appeared on any of the following databases: drug dispensing, hospital discharge diagnoses (the National Minimum Data Set or NMDS), primary care enrolment or laboratory test claims. This population is the ‘health contact population’ for CMDHB in 2007, and forms the denominator for all subsequent calculations. Individuals who had a death recorded in mortality data during this period were removed.

The population was thus limited to those residing within CMDHB, who had appeared in at least one of the previously mentioned data sources in 2007, were aged ≥15 years, and were alive throughout the year. Ethnicity data was taken from 2nd quarter 2007 primary health organisation (PHO) enrolment using a prioritised method in the following order—Māori, Pacific, South Asian and Other. The “Other” ethnicities group was comprised mainly of New Zealand Europeans, with only small numbers of non-European ethnic groups. Following usual practice, we have, therefore, combined these groups under the ‘Other’ category. NZDep2006, an area-based measure of socioeconomic deprivation, was derived from the census area unit recorded in PHO enrolment data.

**Diabetes prevalence using the combined list method**—A diagnosis of diabetes in an individual was determined by a “combined-list” method. That is, an individual was classified as having diabetes if they appeared on at least one of the following three databases:

- Diabetes hospital discharge diagnosis (1998–2007) with ICD 10 codes of E10-E14 or ICD-9 equivalent 250 (diabetes codes), or O24.0 to O24.3 (diabetes in pregnancy) excluding all ICD 10:O24.4 (diabetes arising in pregnancy);
- Three or more HbA1c test claim records during 2006 and 2007 in the laboratory claims database;
- Dispensing of diabetes drugs (May 2005 to 2007) including insulin and all oral hypoglycaemic agents (metformin, glibenclamide, gliclazide, glipizide, and pioglitazone).

The combined list prevalence estimate was then calculated as the number of people with diabetes divided by the health contact population.

**Chronic care management diabetes population**—We investigated the sensitivity of the combined list estimate of prevalence against a population known to have diabetes, from their enrolment in the CMDHB Chronic Care Management (CCMP) programme (taken from 2001 until the end of 2007). To qualify for inclusion in CCMP, individuals must have either Type 1 or Type 2 diabetes with evidence of increased risk for macrovascular outcomes—such as HbA1c ≥9%, blood pressure ≥150/90 mmHg, current smoker, nephropathy (urinary albumin ≥300 mg/L), vision threatening retinopathy, previous cardiovascular disease event, total cholesterol ≥6 mmol/L, or two or more admissions for diabetes to a medical ward for ≥5 bed-days in the last year. Such people are, therefore, more likely to have higher risk of complicated and more advanced disease than people with diabetes who were not enrolled.

**Capture-recapture estimated diabetes population**—A shortcoming of the combined list estimate of prevalence is the possibility that some cases of diabetes may occur but not appear on any of the lists, because they have mild, diet-managed disease; or they have not been diagnosed with diabetes; or because of unreliable recording. In contrast, capture-recapture methods generate an estimated total prevalence without all cases necessarily being recorded in health records. People with diabetes not “caught” or identified on annual lists are estimated by statistical modelling of the degree of overlap between diabetes lists to provide an estimate of the number “not caught”. This number of people with unrecorded diabetes is then added to the number of observed cases to generate the total disease prevalence.

In zoological and epidemiological research, capture-recapture methods may help estimate all ‘cases’ in a population. Biologists initiated these methods to measure the abundance of wildlife populations, by capturing a sample of animals, marking them, and then recapturing a further sample at a later date. An estimate of the total, including the unobserved population, is made by counting the proportion of marked to unmarked animals caught in the second catch. Such techniques have been adapted for epidemiology by using records of health service use as capture occasions. Episodes of “capture” may include hospital discharge diagnoses, drug dispensing, general practitioner diagnosis records, laboratory test use or hospital outpatient diagnosis.

We used the Rcapture utility of the R-project to apply log-linear models to capture-recapture methods. Such models address the problem of dependence between lists that is often encountered in an epidemiological context, where ‘capture’ methods vary and are often closely related to each other. For
example, if a patient is diagnosed with diabetes in hospital, such an event is likely to increase the probability of the subject being treated with diabetes drugs, and consequently, appearing on a second list. In contrast, repeated trapping methods, used to estimate animal abundance, often yield independent samples.

To account for between list dependence, interaction terms, between capture occasions, were included in models. Selecting the model used to assess the total numbers with diabetes in the population involves the trade off between model fit and parsimony. Initially, models with high order interaction terms were computed, with progressively simpler models assessed.

The minimally adequate model, that best fits the data, with a penalty built in for increasing parameters, was judged by comparing Akaike’s Information Criterion (AIC), Chi-square statistics, and plots of Pearson residuals with predicted values for competing models. Scaled rectangle diagrams (like Venn diagrams) display overlap in the datasets used for the combined list and capture-recapture methods and were plotted using SPAN. 7

Capture-recapture information was derived from three separate lists of diabetes patients:

- Pharmaceutical claims for medicines used to treat diabetes;
- Laboratory claims in 1 year with more than 2 HbA1c claims in 1 year; and
- A diagnosis of diabetes in hospital discharge records.

List appearance was limited to the 2007 calendar year, unlike in the combined list analysis, to allow for the possible reversal of diabetes status; to standardise sampling; and to conform to the assumption of a closed population (to simplify model construction).

**Results**

We identified 359,413 people aged ≥15 years living in CMDHB with a health contact in 2007, however, 2437 died in that year, leaving 356,976 individuals for analysis as the denominator population (similar to the 2006 census extrapolated estimate of about 348,000 in this age band). Of this population, 10,600 (3.0%) had a diagnosis of diabetes recorded in the last 9 years in hospital records. 17,991 (5.0%) had three or more HbA1c tests recorded in the 2-year interval between 1 Jan 2006 and 31 Dec 2007; and 21,633 (6.1%) had at least one drug dispensing for diabetes-related medication between May 2005 and 31 Dec 2007.

Overall (see Table 1), the combined list estimate of diabetes prevalence was 7.2% (25,797/356,976); compared with capture-recapture estimates, which were about 30% higher at 9.5% (34,418/356,976). Marked variation occurred in the prevalence of diabetes by age and ethnic group for both the combined list and capture-recapture estimates. In middle-age (35 to 54) the combined list frequency of diabetes is between 1 in 20 and 1 in 10, then it increases to between one in seven to 1 in 5 people in the 55 to 75 age group.

Of the ethnic groups, Pacific and South Asian people had the highest prevalence by both estimation methods. South Asian people show the largest difference between the two estimation methods (12.3% by combined list compared to 20.5% by capture-recapture); but the precision of the South Asian capture-recapture prevalence is poor due to small numbers. There was no evidence of an increasing socioeconomic status gap between the combined list and capture-recapture estimates.

Of those in the denominator (health contact) population, 10,725 individuals (Table 1) were enrolled in the diabetes CCM module until the end of 2007. The combined list correctly identified (sensitivity) 10,351/10,725 (96.5%) of these people.
Considerable overlap was observed between the three diabetes indicators, and those known to have diabetes in the CCMP diabetes register (Figure 1). The areas of the three coloured boxes represent the size of each of the three lists used to define diabetes status, while the numbers relate to the mutually exclusive non-overlapping or overlapping categories within the combined lists. Most people were identified from drug dispensing records.

Table 1. Sociodemographic characteristics of people with diabetes: i. ‘diagnosed’ by the Combined-list (algorithm) method; ii. estimated by the Capture-recapture method; iii. enrolled in the Chronic Care Management (CCMP) diabetes programme: presented as numbers (and percentages) of the 2007 Counties Manukau population aged 15 years and over (n=359,413)

<table>
<thead>
<tr>
<th>Variables</th>
<th>i. Combined list estimates of diabetes prevalence: n (%)‡‡</th>
<th>ii. Capture-recapture estimates of diabetes prevalence: % (95% CIs) ††</th>
<th>iii. Patients with diabetes enrolled in the CCMP: n (%) ‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>13,223 (7.0)</td>
<td>9.3 (8.6 to 10.2)</td>
<td>5491 (2.9)</td>
</tr>
<tr>
<td>Male</td>
<td>12,572 (7.4)</td>
<td>10.0 (9.3 to 11.1)</td>
<td>5234 (3.1)</td>
</tr>
<tr>
<td>Age category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 to 24</td>
<td>496 (0.7)</td>
<td>1.0 (0.24 to 1.8)</td>
<td>106 (0.1)</td>
</tr>
<tr>
<td>25 to 34</td>
<td>1221 (1.9)</td>
<td>2.7 (2.0 to 5.4)</td>
<td>360 (0.6)</td>
</tr>
<tr>
<td>35 to 44</td>
<td>3258 (4.5)</td>
<td>7.3 (5.7 to 11.2)</td>
<td>1318 (1.8)</td>
</tr>
<tr>
<td>45 to 54</td>
<td>5487 (9.2)</td>
<td>12.7 (11.0 to 15.9)</td>
<td>2446 (4.1)</td>
</tr>
<tr>
<td>55 to 64</td>
<td>6863 (15.9)</td>
<td>20.0 (18.3 to 22.7)</td>
<td>2955 (6.8)</td>
</tr>
<tr>
<td>65 to 74</td>
<td>5264 (20.6)</td>
<td>27.2 (24.5 to 31.4)</td>
<td>2289 (9.0)</td>
</tr>
<tr>
<td>over 75</td>
<td>3208 (17.8)</td>
<td>22.3 (20.2 to 25.5)</td>
<td>1251 (6.9)</td>
</tr>
<tr>
<td>Ethnic group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Māori</td>
<td>3310 (7.6)</td>
<td>8.9 (8.2 to 10.0)</td>
<td>1737 (4.0)</td>
</tr>
<tr>
<td>Pacific</td>
<td>7661 (11.1)</td>
<td>13.7 (12.6 to 15.4)</td>
<td>4013 (5.8)</td>
</tr>
<tr>
<td>South Asian</td>
<td>2510 (12.3)</td>
<td>20.5 (15.3 to 36.9)</td>
<td>917 (4.5)</td>
</tr>
<tr>
<td>Other†</td>
<td>12,316 (5.5)</td>
<td>7.6 (7.0 to 8.6)</td>
<td>4058 (1.8)</td>
</tr>
<tr>
<td>NZDep2006*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 and 2 least deprived</td>
<td>3413 (5.1)</td>
<td>7.3 (6.1 to 9.7)</td>
<td>937 (1.4)</td>
</tr>
<tr>
<td>3 and 4</td>
<td>2668 (5.2)</td>
<td>7.6 (5.3 to 9.9)</td>
<td>785 (1.5)</td>
</tr>
<tr>
<td>5 and 6</td>
<td>2023 (5.5)</td>
<td>7.3 (6.1 to 9.9)</td>
<td>612 (1.7)</td>
</tr>
<tr>
<td>7 and 8</td>
<td>4187 (8.0)</td>
<td>8.0 (6.6 to 11.1)</td>
<td>1741 (3.3)</td>
</tr>
<tr>
<td>9 and 10 most deprived</td>
<td>13,501 (9.0)</td>
<td>11.1 (9.6 to 13.6)</td>
<td>6650 (4.5)</td>
</tr>
<tr>
<td>Total</td>
<td>25,797 (7.2)</td>
<td>9.6 (9.1 to 10.3)</td>
<td>10,725 (2.3)</td>
</tr>
</tbody>
</table>

*Includes 70 persons with unknown gender and 78 persons with unknown NZDep2006 status; †Other is mainly European ethnic group; *NZDep2006 – area based measure of socioeconomic deprivation; ‡CCMP is the Chronic Care Management Programme for people with diabetes who are considered at high risk of re-admission to hospital; ††As this method provides a modelled assessment of people with diabetes we have presented prevalence information, rather than absolute numbers; ‡‡CIs not presented for Combined list prevalence estimates as all are less than 0.5%.
The number of people assumed to have diabetes in the three databases used for the combined list estimates were also used to derive the overall capture-recapture estimates of diabetes prevalence. As the data for the latter analyses were restricted to 2007 only (discussed in the Methods section), we used two or more HbA1c tests in 1 year as laboratory evidence of diabetes, compared with the combined list method, in which three tests in 2 years as laboratory evidence of diabetes. The model fit for the 2007 capture-recapture estimate of diabetes prevalence was optimal using a log-linear model that included two-way interactions for all lists. The best fit (indicated by lowest AIC - see Methods - with least discordance between observed and expected values and plots of residuals and predicted values) was 34,418 (95% CI 32,523 to 36,812). This translates to an estimated diabetes prevalence of 9.6%, or about 8,600 more cases compared to the combined list method. If the
capture-recapture results (point estimate) were assumed to be the best estimate of the number of people with diabetes in this population, then the combined list method would have identified 75% (25,797/34,418) of all people aged 15 years and over with diabetes in the CMDHB.

**Discussion**

Our method for estimating the prevalence of diabetes, that combines linked routine health databases, provides a reasonably accurate estimate of known diabetes prevalence. Such a method shows good concordance with capture-recapture estimates and a list of people enrolled in a diabetes management programme. Local surveys have reported diabetes prevalence consistent with our combined-list estimates. For example, the New Zealand Health Survey 2006/7, reported an adult (age ≥15 years) prevalence of diagnosed diabetes in CMDHB of 26,400, very similar to our combined-list prevalence (25,797).

As the combined list method can only identify people known to have diabetes – because they have to be recorded in a routine health dataset for a ‘diabetes-related’ activity - it will miss undiagnosed diabetes. This shortcoming is illustrated by the higher estimated prevalence using the capture-recapture method, that is able to estimate the size of this unidentified group. When compared to the capture-recapture estimate, the undercount using the combined list approach was about 25%, or 8600 people with diabetes.

The gap we observed between the two methods of estimating diabetes prevalence is consistent with an earlier (2002/03) community survey,\(^8\) undertaken in the same geographic region, that reported an undercount of 33% in European, 25% for Māori and 17% for Pacific, and a weighted mean undercount of about 28%. Taken together, these estimates suggest that undiagnosed diabetes rates have been relatively constant in the interval years, despite effort to improve rates of diabetes detection.\(^9\)

Gaps between the combined list and capture-recapture prevalence may help identify subgroups of the population with high levels of undiagnosed diabetes. The largest gaps observed between the two prevalence estimates were for South Asian people although the precision of the capture-recapture estimates were poor. To better assess such gaps, we plan to combine data from several years or use larger populations to generate the necessary statistical power to investigate this observation further. However, if our estimates were correct, a significant population of about 1,700 South Asian people in CMDHB, had diabetes, but were not diagnosed during 2007.

The interpretation of the apparent differences in prevalence between the two methods is limited by the validity of capture-recapture methods, which is contingent on several underlying assumptions, we introduced earlier. One assumption, of a closed population, is approximated here by limiting records of health care use to one year. Another includes between-list independence, which is questionable in an epidemiological context, as the lists used in the modelling are, frequently, not independent. We addressed this problem with log-linear models, that has been discussed at length in another summary.\(^3\) In short, such models are designed to account for between list dependence through the inclusion of interaction terms, with our final model including beta coefficients for all two-way dependence between the three lists. With three lists used in the capture-recapture, this level of dependence was
the maximum possible, given that one interaction term (three-way in our study) must be dropped, to solve the equation. Capture-recapture models are also dependent on the assumption that counts of people with diabetes on lists, and between-list intersections, are distributed according to a Poisson probability model. Therefore, the capture-recapture prevalence we calculated should be interpreted with considerable caution, and we do not suggest using such results as the best estimate of diabetes prevalence.

The combined list method will produce more conservative estimates of diabetes prevalence and some people with mild diabetes, managed by diet alone, are likely to be missed. We compared the combined list method to a gold standard of patients with diabetes enrolled in a chronic care management programme. CCMP patients represent a subgroup of people with diabetes with increased risk of complications given the eligibility criteria for the programme. Using this standard, our combined list method performed well, which is not surprising because, as discussed above, the combined list method is more likely to identify people with more longstanding and complicated disease.

The use of historical data, and differing time periods for each dataset (using as much data as was available), could potentially lead to mismatches, for example through out-migration. We controlled for this by ensuring that individuals were all present in CMDHB in 2007 for at least one health contact. Also, other work has shown that as the health datasets improve in quality over time, the health contact population begins to approximate Census-based estimates.

Both methods prevalence estimates assume that the diabetes-related activities recorded in the three databases accurately identify people with the disease. The weakest indicator of a diabetes diagnosis is likely to be that based on HbA1c test records. These may be requested (repeatedly) in patients suspected, but not fulfilling, the diagnostic criteria for diabetes. Although not recommended as a screening test in national guidelines, HbA1c tests are increasingly used for this purpose. When used to monitor glycaemic control, protocols recommend a 3 to 6 monthly testing interval. We used three HbA1c tests in 2 years to define diabetes for the combined list estimate, based on a receiver operating characteristic curve, which used CCMP status as the ‘gold standard’ comparator (Figure 2). The curve shows that optimal performance (assuming the relative costs of false positives to false negatives are identical) is achieved at a level of three, per 2-year period. Such performance may change as HbA1c testing becomes more frequently used to screen for diabetes status (already recommended in the US). Of the people identified with diabetes using the combined list, 10.5% (2,714/25,797) were defined as a result of HbA1c test records only. An arguably, stricter definition of diabetes, based on two HbA1c measures in 1 year rather than three in 3 years, was applied to the capture-recapture definition, as the databases used were all restricted to 1 year.
Internationally, diabetes prevalence studies have reported both combined list and capture-recapture methods, although not together. The combined list estimate method has been used previously in Denmark\textsuperscript{12} to describe time trends in diabetes prevalence between 1995 and 2006. The researchers used a similar technique to the one we describe; however, five or more blood glucose measurements in a year, rather than three HbA1c measures in 2 years, were used as laboratory evidence of diabetes. They also included diabetes outpatients visits as one of their lists. A similar prevalence study from Ontario was based on a rule that included a diabetes diagnosis in hospital discharge or outpatient records in the last 2 years.\textsuperscript{13} The sensitivity of 96.5\% for the combined list method in our study is higher than that reported in the Danish (85\%)\textsuperscript{12} and Canadian (86\%) studies,\textsuperscript{14} but, as discussed, this may be overestimated due to the
severity of disease in our gold standard group of patients in the CCMP. Our capture-recapture result suggests the combined list method has a sensitivity of about 75% for all diabetes.

The Casale Monferrato study in northwest Italy used capture-recapture log-linear models, similar to ours, to monitor the incidence of diabetes between 1988 and 2000. Data from diabetes clinics, hospital discharge records, prescribing and sales of reagents and strips were combined to calculate population estimates. Comparing individual to capture-recapture methods, they found a case-ascertainment of about 80%; which was consistent with our estimate of 75%. In the United Kingdom, similar rates of list undercount to the Italian study have been reported in capture-recapture studies of diabetes prevalence.

Our study shows that diabetes prevalence, derived from combining linked health databases, is similar to survey-based estimates in South Auckland. Further, the discrepancy between combined list and capture-recapture diabetes prevalence estimates appears greatest among people who identify as South Asian. In this area, primary care screening programmes for diabetes are currently targeting Pacific and Māori people, yet the greatest proportion of undiagnosed diabetes may now exist in the South Asian population.

Given the small numbers of people in many ethnic subgroups in our study, we had insufficient power to investigate these groups adequately, and so plan to repeat the current study in larger populations. In addition, the incomplete nature of the CCMP diabetes list did not allow us to estimate the specificity of the combined list method. We therefore plan to examine this issue with a more complete “gold standard” list of people with diagnosed diabetes drawn from a large, primary care database.

Competing interests: None.

Author information: Simon Thornley, Research Fellow, Section of Epidemiology and Biostatistics, School of Population Health, Tamaki Campus, University of Auckland; Roger Marshall, Associate Professor of Biostatistics, Section of Epidemiology and Biostatistics, School of Population Health, Tamaki Campus, University of Auckland; Gary Jackson, Public Health Physician, Counties Manukau District Health Board, Manukau City; James Smith, Public Health Medicine Specialist, Auckland Regional Public Health Service, Greenlane Clinical Centre, Auckland; Wing-Cheuk Chan, Public Health Medicine Specialist, Counties Manukau District Health Board, Manukau City; Craig Wright, Senior Advisor, Ministry of Health, Wellington; Dudley Gentles, Biostatistician, Section of Epidemiology and Biostatistics, School of Population Health, Tamaki Campus, University of Auckland; Rod Jackson, Professor of Epidemiology, Section of Epidemiology and Biostatistics, School of Population Health, Tamaki Campus, University of Auckland

Acknowledgements: This work was partially funded by a grant from the New Zealand Ministry of Health. We also thank Dean Papa for assistance with data extraction and management.

Correspondence: Simon Thornley, Section of Epidemiology and Biostatistics, School of Population Health, Tamaki Campus, University of Auckland, Auckland, New Zealand. Fax: +64 (0)9 2629501; email: s.thornley@auckland.ac.nz
References:


