Public Health Monograph Series
No. *39*
ISSN *1178-7139*

Technical Report for BODE^3 Pre-diabetes Multistate Lifetable Model

Version 1

Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme (BODE^3)

Technical Report: Number 20

Anna Davies
Tony Blakely
Cristina Cleghorn

Version 1: April 2022

A technical report published by the Department of Public Health,
University of Otago, Wellington

ISBN 978-0-9951070-4-5

Contact: Cristina Cleghorn
Dietary modelling lead in BODE^3 Programme, University of Otago, Wellington, New Zealand.
Email: Cristina.cleghorn@otago.ac.nz

Contents

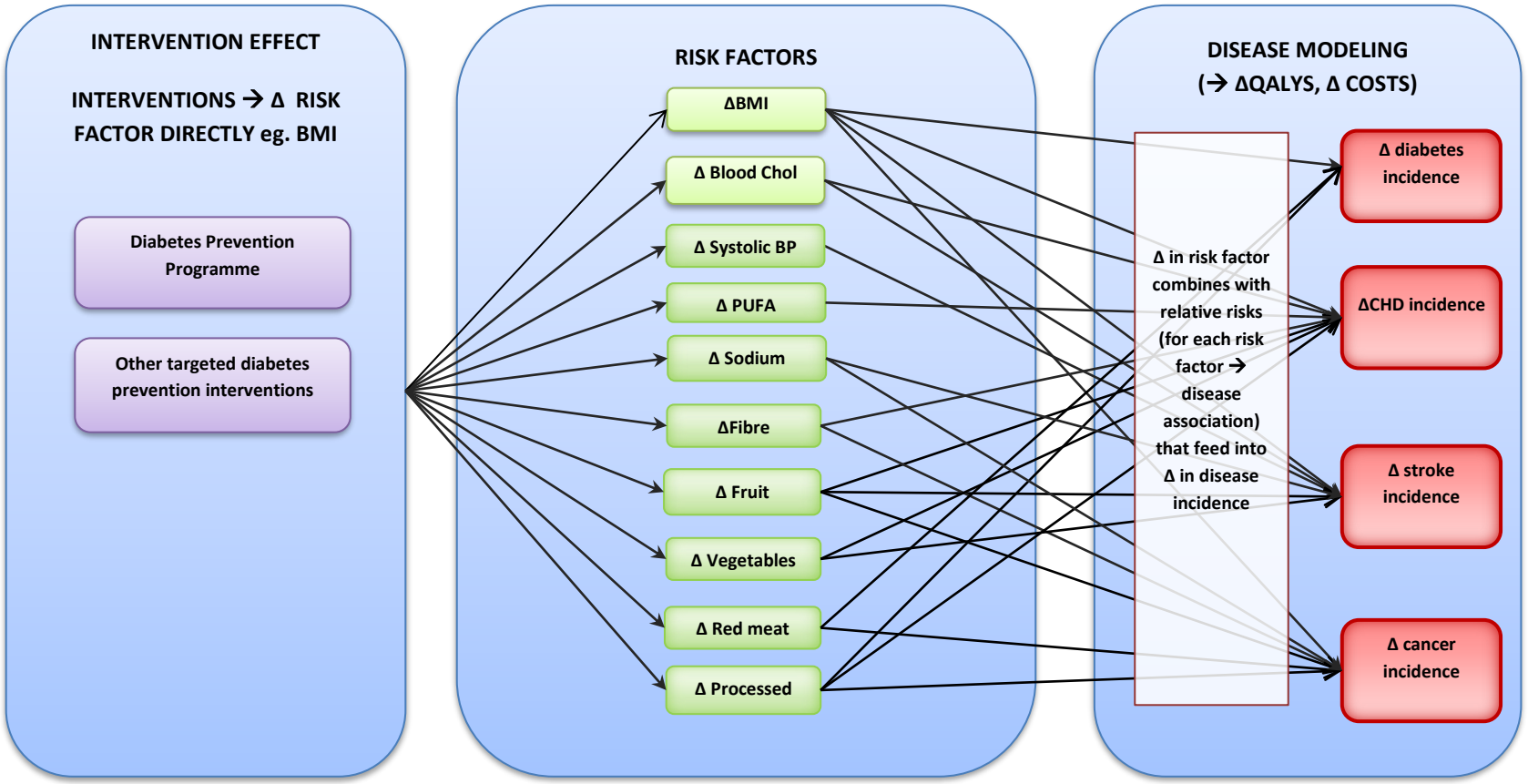
Overview and purpose

This Technical Report provides the documentation on the Burden of Disease Epidemiology Equity and Cost Effectiveness ($BODE^3$) pre-diabetes multistate lifetable (MSLT) model. This model estimates the effect the change in risk factor has on *health impacts* and *cost impacts* of a range of interventions in the New Zealand pre-diabetes population, with the ability to examine *heterogeneity by sex, age and ethnicity*.

This model is a variation of the DIET MSLT model and detailed methods on the development of the DIET MSLT model can be found in the 'Technical Report for $BODE^3$ Diet Intervention and Multistate Lifetable Models'.[3] This technical report will detail the differences between the pre-diabetes MSLT model and the DIET MSLT model.

The conceptual structure of the pre-diabetes MSLT model is shown below in Figure 1. Specific dietary interventions lead to change in dietary risk factors that are entered as inputs into the pre-diabetes MSLT model. These changes lead to changes in disease incidence. The dietary and physiological risk factors included in version 1 (V1) of the pre-diabetes MSLT model are BMI, systolic blood pressure (BP), total blood cholesterol, sodium, polyunsaturated (PUFA) fat, fibre, fruit, vegetables, red meat and processed meat intake. All these risk factors have epidemiological associations with diseases as reported in the Global Burden of Disease (GBD)(4). All the diseases included in the DIET MSLT have been included in the pre-diabetes MSLT model to allow for any of the DIET risk factors to be included.

Figure 1: Conceptual model for dietary interventions



Pre-diabetes population numbers

The pre-diabetes MSLT model includes estimates of the number of people with pre-diabetes by age, sex and ethnic group. The NZ national adult nutrition survey 2008/09 was used to calculate the percentage of each age, sex, ethnic group that was pre-diabetic in this representative survey. These percentages were then applied to the total population numbers in the whole population DIET MSLT model to estimate the number of people with pre-diabetes in each age, sex and ethnic group. These numbers were used to populate the pre-diabetes MSLT model. Those that were pre-diabetic were defined as having HbA1c values between 41 and 49.

Risk factors and diseases in the pre-diabetes MSLT model

In the first version of the pre-diabetes MSLT model, developed for the Food for Health HRC project and the BODE^3 HRC programme, the following risk factors are included with associations onto the following diseases:

- BMI (14): CHD, stroke, diabetes, osteoarthritis, cancers: endometrial, kidney, liver, oesophageal, pancreas, thyroid, colorectal, breast, ovarian, gallbladder
- Systolic blood pressure (1): Stroke
- Total blood cholesterol (2): CHD, stroke
- Sodium (3): CHD, stroke, stomach cancer
- PUFA (1): CHD
- Fibre (2): CHD, colorectal cancer
- Fruits (5): CHD, stroke, cancers: head and neck, lung, oesophageal
- Vegetables (2): CHD, stroke
- Red meat (2): Diabetes, colorectal cancer
- Processed meat (3): CHD, diabetes, colorectal cancer

Specification of pre-diabetes cohort used to populate model disease rates

Purpose of the pre-diabetes cohort

To adapt the DIET MSLT model to function as a model for a pre-diabetic population, rates of disease amongst the pre-diabetic population needed to be calculated. These rates were used to replace disease rates for the whole population used in the DIET MSLT model, where the rates in the two populations are known to differ significantly, based on evidence from literature.

In order to determine rates of disease specific to people with pre-diabetes, we created an algorithm which flags whether or not someone has interacted with the publicly-funded health system in a way which indicates they may have pre-diabetes, to create a "pre-diabetes cohort". The algorithm is based on data from the national health collections [(5)] and works at the encrypted National Health Index (NHI) [(6)] level – so while it flags individual people, we can't tell who they are (name, address etc). Because the algorithm works at the person level, we can then examine some of the characteristics of people with pre-diabetes and rates of disease in this group, such as prevalence, incidence and mortality rates for coronary heart disease, stroke, stomach cancer, endometrial cancer, and pancreatic cancer, using data readily available in the national collections.

This section will outline how we developed the algorithm we use to define the pre-diabetes cohort, covering these key questions:

- What data sources were available to identify people with pre-diabetes?
- What are some of the issues associated with these sources?
- What algorithm options did we try, and what methods did we use to test them?
- What did we find: which algorithm worked best?
- What did we conclude about which algorithm to use, and what do we need to keep in mind when using it?

What data sources were available to identify people with pre-diabetes?

The definition of pre-diabetes varies across the world. New Zealand generally uses the HbA1c test to screen for diabetes and pre-diabetes, with or without other tests / presence of symptoms (see diabetes screening guidelines in Appendix A, New Zealand Society for the Study of Diabetes [(7)]). An HbA1c test result of 41 to 49 mmol/mol is considered to indicate pre-diabetes.

To create an algorithm that flags whether or not someone may have pre-diabetes, we needed to understand how someone with pre-diabetes might use the health system – and at what point data may be recorded about their pre-diabetes status and service use (and subsequently reported to national health collections). We identified two main points where this may happen: laboratory testing, and hospitalisation events.

ICD coding of pre-diabetes, and hospitalisation data

Some people with pre-diabetes may spend time in hospital (for a range of reasons), and their pre-diabetes status could be recorded as part of the diagnosis information for hospital discharges reported to the NMDS.

The coding of pre-diabetes in the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD-AM-10) [(1)] changes slightly depending on the edition. In the First and Second editions, the R73 code ("Elevated blood glucose level") splits out into: R730 - Abnormal carbohydrate tolerance, and R739 – Hyperglycaemia, unspecified. Pre-diabetes is listed as one of the conditions under the R730 code – other conditions listed under this code are : Abnormal glucose tolerance, Diabetes – chemical, Diabetes – latent, Impaired glucose tolerance.

From the Third edition onwards, the coding doesn't split out beyond R73 (Elevated blood glucose level). But the Third, Sixth and Eigth editions all contain E09 codes ("Impaired glucose regulation" in the Third and Sixth editions and "Intermediate hyperglycaemia" in the Eigth edition – which are described as synonyms for pre-diabetes in the coding notes). The E09 codes map back to R730 when you map backwards to version two, and R730 maps into E09 going forwards from version two to version three.

Hospital diagnosis data in the NMDS is coded in, or can be mapped to, ICD-10-AM third edition, so we can identify individuals (anonymously) with a diagnosis of E09 (clinical code system 12), indicating that they have pre-diabetes.

Information from the National Minimum Dataset (NMDS) [(8)] containing diagnosis data reported in or mapped to ICD-10-AM version 3 code E09 can be used to (anonymously) identify people with pre-diabetes who have been discharged from hospital, focusing on publicly-funded events only. This will include only some of the people with pre-diabetes in New Zealand – it will not identify people with pre-diabetes who have not had a publicly funded hospital discharge.

The second data source for the pre-diabetes algorithm is data from laboratory tests. The ideal data would be HbA1c test results by person for all of New Zealand – but nationally collected claims data (Laboratory Claims Collection [(9)]) does not contain test results – just the fact the an HbA1c test took place on a particular date for a particular person. TestSafe data [(10)] has lab test results but only for northern region DHBs, not the whole of NZ, and we were unable to access to TestSafe data for this work (the data extract costs were prohibitive).

Most people with <u>known</u> pre-diabetes should be having regular HbA1c testing. For example, Braatvedt et al. [(11)] note that "for patients with an initial HbA1c result of 41–49 mmol/mol, cardiovascular risk assessment and lifestyle interventions are recommended with repeat HbA1c screening in 6–12 months" (p. 70). These recommendations are drawn from the screening guidelines published by New Zealand Society for the Study of Diabetes in 2011 (NZSSD -[(7)], See Appendix A).

More recent advice on testing recommendations from the New Zealand Ministry of Health in 2016 contained in "Pre-diabetes: risk factor management" indicate a similar testing frequency: "Initial

HbA1c should be repeated after three months of lifestyle change and thereafter at six to twelve month intervals" [12].

If New Zealand physicians are generally following the guidelines around HbA1c testing for people with pre-diabetes, it may be possible to use the patterns of HbA1c testing as an indication of pre-diabetes, in the absence of the test results themselves. Using lab test patterns from the Labs collection is part of algorithms used to indicate other health conditions – for example: diabetes. The Virtual Diabetes Register (VDR) [13] uses a number of rules to indicate whether an individual is likely to have diabetes, including patterns of lab testing (HbA1c and albumin/ creatine ratio tests).

Virtual Diabetes Register (VDR) algorithm

Inclusions

- Hospital discharges coded for Diabetes Mellitus (DM) in the previous ten years
- Outpatient attendances for DM education management and DM retinal screening in the past three years
- Subsided community dispensings of pharmaceutical therapies typically used by people with diabetes, over the most recent two year period
- Laboratory tests ordered where there have been four or more HbA1c tests and two or more urinary albumin/creatinine ratio (ACR) tests over the most recent two-year period.

Exclusions

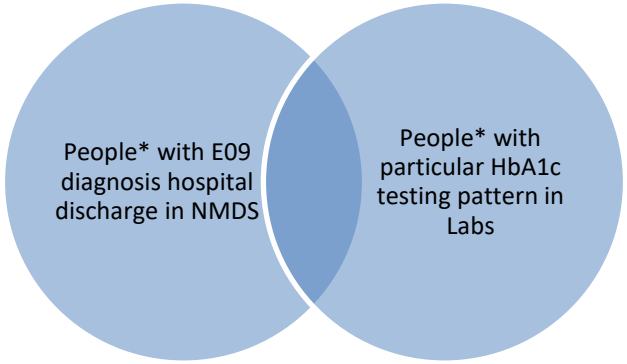
- Hospital discharges coded for gestational diabetes
- Metformin dispensings for women aged 12-45 (Metformin is used to treat Polycystic Ovary Syndrome and diabetes)
- Some lab test and pharms information is not included for women who have given birth (claims around the time of birth).

With this in mind, we created a pre-diabetes algorithm for health system users based on these two key data sources:

- People who have a publicly funded hospital discharge diagnosis E09, as reported to the NMDS
- People who have a particular pattern of HbA1c testing, as reported in the Laboratory Claims collection.

We will exclude people with diabetes (defined using the VDR criteria) and the HbA1c tests for women during pregnancy (as these tests could be for women with gestational diabetes).

Figure 2: Main data sources for pre-diabetes algorithm



*Excluding people with diabetes (using VDR definition), and tests for women receiving HbA1c testing during pregnancy

What are potential issues associated with these data sources?

Before creating the pre-diabetes algorithm, we investigated both data quality issues associated with the data collections, and the potential pitfalls associated with using patterns in lab testing as an indicator of pre-diabetes status. Generally, the NMDS is considered to be a stable collection, and diagnosis information of good quality. The potential use of HbA1c test information from the Laboratory Claims collection required a more detailed examination.

Bulking funding of laboratory tests

No national collection of health data is without issues, and the Laboratory Claims collection is no exception. A key issue relates to how laboratory tests are funded. The way tests are funded has changed over time from a fee-for-service payment to bulk funding. The move towards bulk funding means that there are fewer incentives for providers to submit accurate individual claims, and there a fewer incentives for the Ministry of Health to prioritise data quality initiatives for the Laboratory Claims Collection (such as updating business rules to upload claim information - last done in 2010). It is likely that some of the data in the collection is inaccurate, and that some laboratory claim data is missing. As part of investigating how to create an algorithm for pre-diabetes using lab tests, an appraisal of the data quality specific to the HbA1C test was conducted.

A key data quality issue affecting the collection is that new tests introduced after 2010 are not part of the collection. However, this does not affect the HbA1C test, claims for which pre-date 2010.

Potential impact on pre-diabetes algorithm: low to medium. The HbA1c test pre-dates the 2010 period, so is included in the collection. However, bulk-funding may affect HbA1c test volumes in the collection so this should be investigated.

Laboratory Claims NHI coverage

The proprotion of claims with valid NHIs may have changed over time. In earlier years, many claims had no NHI. This could may mean that testing frequency is not accurately captured for a person (e.g. if some tests can not be traced back to an individual). This was more a problem for earlier years when the collection first began and less of a problem for more recent periods. According to the Ministry of Health: "In 2010 claims, the encrypted NHI number is stored for approximately 98 percent of laboratory test records. (In earlier years, it varied, dropping to as low as 13 percent in 1997 claims.)" [(9)]

Potential impact on algorithm: low. NHI coverage is good in recent years, which are the periods of interests for this work.

Changes over time for HbA1c testing – related to health targets

Laboratory testing patterns have changed over time – one notable pattern is the increase in HbA1c testing generally observed from around 2011 and 2012. During this time there were policy and measurement changes relating to HbA1c testing occuring in New Zealand.

During 2011, the way HbA1c was reported changed from percent to mmol/mol, in line with international movements (originating from the International Federation of Clinical Chemistry) (Braavedt et al. [(11)]). In late 2011, the New Zealand Society for the Study of Diabetes (NZSSD) introduced guidelines for using HbA1c tests in diabetes screening (using the new mmol/mol units), and for clinical pathways to follow based on HbA1c results (including the use of the HbA1C test for monitoring).

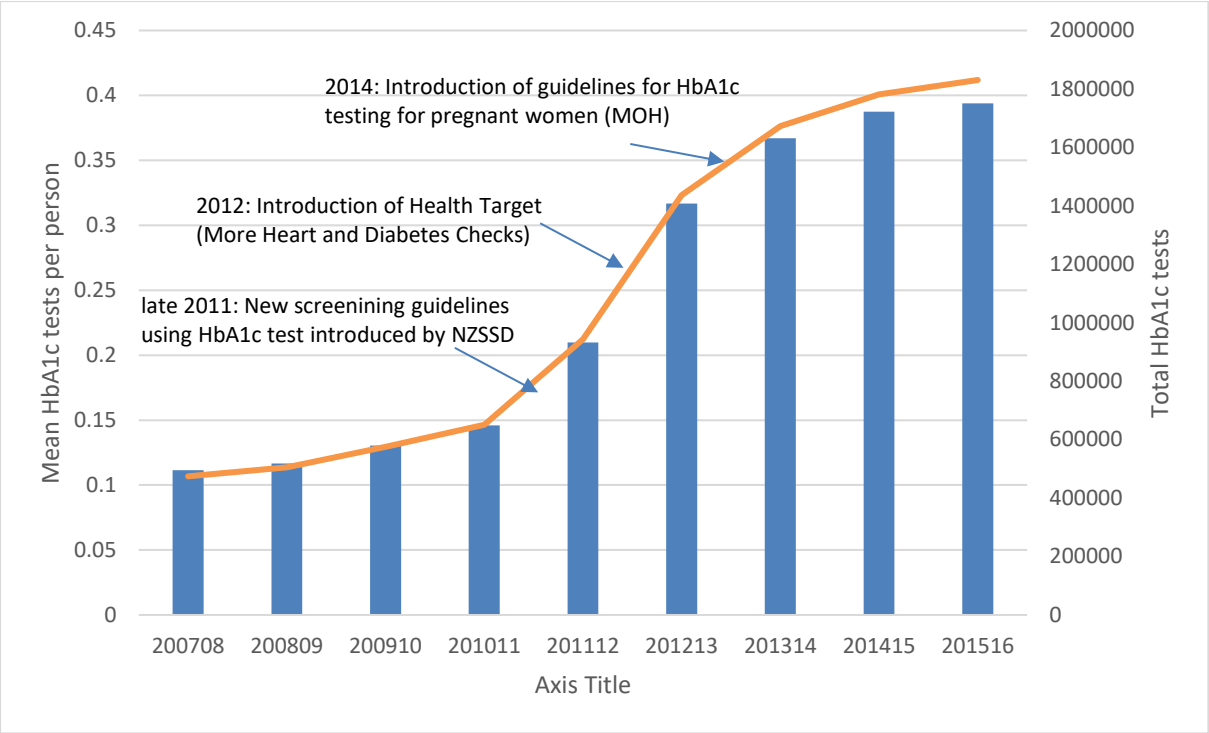
Along with the changes in HbA1c measurement and screening guidelines, around this time the government introduced a health target called *More Health and Diabetes Checks* which provided financial incentives for DHBs to achieve the target.

"The *Checks* health target has been operating since 2012 and includes a cardiovascular risk assessment (CVDRA) and a blood test for diabetes (HbA1c) delivered in primary care settings. The goal of the health target was for 90 percent of people in specified age and ethnicity cohorts to have had a CVDRA in the past five years. The *Checks* health target budget included national funding to support the target, and incentives and sanctions for district health boards (DHBs) and primary health organisations (PHOs) to achieve the target. The funding pool was spread over fiscal years 2013/14; 2014/15; 2015/16; and 2016/17. The amount available reduced each year." (Allen & Clark, 2016, "More Hearts and Diabetes Checks Evaluation" [(14)])

Advice around HbA1c testing in pregnancy has also changed over time. In December 2014, MOH released guidelines for diabetes testing in pregnant women [(15)]. The guidelines noted that screening and diagnosis of diabetes in pregnant women varied across different parts of the country. The recommendations now include conducting an HbA1c test as part of the tests undertaken before 20 weeks gestation.

These events are noted in Figure 3 below, which shows the total number of HbA1c tests across NZ (orange line), and average number of HbA1c tests per healthcare user over time (blue columns).

Figure 3: Mean HbA1c tests per person (columns) and total HbA1c tests (line) amongst health service user population



Potential impact on pre-diabetes algorithm: medium to high. The changing patterns of HbA1C testing over time will mean that, in the creation of a pre-diabetes indicator based on lab test frequency, results for years prior to 2012 will be probably be quite different to results for the 2012 year and beyond. Investigation into reporting voulmes over time is required.

<u>Frequency of testing only partially related to pre-diabetes status?</u>

It could be that the frequency of HbA1C testing is only partially related to pre-diabetes status. That is, NZ physicians could be departing from the diabetes / pre-diabetes screening guidelines in practice, or frequently screening healthy people.

Because we do not have the results of the HbA1C tests, we cannot directly investigate the extent to which patterns in HbA1C testing reflect pre-diabetes status. However, as part of developing the pre-diabetes algorithm we can test the algorithm results against other sources of information. For example, we can compare pre-diabetes prevalence generated based on the pre-diabetes algorithm with prevalence found amongst the New Zealand population through the Adult Nutrition Survey (e.g. as reported in Coppell et al. [16]), which was based on HbA1c test results. Similar patterns in prevalence by age, sex, and ethnic group for similar time periods, may indicate that the algorithm is identifying the right groups of people.

Similarly, we can also look at rates of disease amongst the pre-diabetes algorithm cohort and compare these patterns with published evidence from other sources. For example, we can compare

the development of diabetes in the pre-diabetes algorithm cohort with the development of diabetes in people with pre-diabetes from one specific New Zealand primary care organisation (Teng et al. [17])

It also may be possible that the testing patterns are more closely aligned to the diabetes screening guidelines in particular years, and less well-aligned in other years, so looking at the pattern over time may be useful as well.

Potential impact on pre-diabetes algorithm: medium to high. We have no direct way to test this with the data we have. However, comparing the pre-diabetes prevalence / rates of disease generated by the algorithm with other sources of the same or similar information will provide us with an assessment of how well the indicator picks up people with pre-diabetes (and an indirect way to estimate how well test frequency is related to true pre-diabetes status).

What algorithm options did we try, and what methods did we use to test them?

Testing and development aims

In general, with testing any algorithm of this nature, it is useful to keep in mind the purpose for the algorithm. In this case, the main purpose for developing the algorithm is so that we can look at disease rates amongst people with pre-diabetes. The aim in developing the pre-diabetes algorithm was not to develop a predictive tool, or to be able to create a disease register of people with pre-diabetes.

Disease register
"A disease register is a documentation of all cases of a certain disease or health condition, which occur within a defined population. Registers are held by registries, which are the systems in place for the continuous registration of cases. In general, clinical sources prospectively notify new cases as they arise in the clinical setting. Along with personal information, clinical data that is routinely collected in clinical practice, is submitted, with the notified variables differing by registry and disease of interest." [2]

At the most basic level, we needed to understand who the algorithm picks up, and who it doesn't. We can work this out by comparing pre-diabetes prevalence and incidence rates for the algorithm cohort with similar kinds of rates from other sources.

Specific to the purpose for which the algorithm is developed, we need to consider whether there any aspects of this group that might mean they are not representative of the whole pre-diabetes group and how this may affect rates of disease based on the algorithm cohort. The nature of the data sources used for the algorithm mean that the algorithm can only capture known cases of pre-diabetes, with some level of health system contact.

Which algorithms to try?

The guidelines from the NZSSD recommend that those with pre-diabetes (with an HbA1c score in the 41 to 49 mmol/mol range) are tested every 6-12 months. It is easier to frame this in terms of testing

frequency over a two year time period, as it gives a bit more room to capture testing done annually (if some years the "annual" testing gets done every thirteen months some of the time, for example).

Therefore, if a doctor were following the HbA1c testing guidelines for someone with (diagnosed) pre-diabetes, HbA1c screening would probably happen about 2-4 times in two years.

- The minimum would be two tests in a two-year period (i.e. testing every twelve months).
- For people who were more closely managing their pre-diabetes – testing every six months, the test count in a two year period might be 4 tests
- Testing every three months would equate to eight tests in a two-year period, roughly.

In terms of an upper limit of useful HbA1c testing, it may be possible to look at testing every 10 weeks, given that the HbA1c test "measures your average blood glucose over the previous 8 to 12 weeks" (Health Navigator [18]) – ten weeks being the midpoint of 8 to 12. In terms of a two-year period, this would be about 10 tests in two years (104 weeks / 10 = 10.4)

Four or more HbA1c tests in a two year period, plus 2 or more urinary albumin/creatinine ratio (ACR) tests, would indicate diabetes (in line with the VDR algorithm), so a pre-diabetes algorithm could look at testing patterns occurring less frequently that the diabetes algorithm rules – which gives a range of 2 -3 tests over a two year period.

We examined four different options for HbA1c testing patterns as part of the pre-diabetes algorithm:

Table 1: Four algorithm options for HbA1c testing patterns

Number of HbA1c tests in a two year period	Rationale:
2-3	The number of HbA1c tests sits below the VDR threshold
2-4	Likely testing pattern range if following the pre-diabetes HbA1c testing guidelines (MOH or NZSSD)
2-10	Minimum level according to guidelines would be 2 tests; upper limit bounded by average amount of time the test actually covers (ten weeks)
2 or more (no upper limit)	Using the minimum level of testing only.

These people may or may not also have a publicly funded hospital discharge of E09.

Along with the exclusion of people with diabetes (occurring within or before the period of interest for capturing people with pre-diabetes), we would also apply an exclusion around testing for pregnant women (as the VDR algorithm does). We will exclude all HbA1C lab tests for pregnant women (date range identified as event end date and nine months before event end date, for births events).

Pre-diabetes "diagnosis date"

Another decision is what date to use to as a "diagnosis date" for the labs data, given it is based on a sequence of visits. Although, we cannot exactly determine the date of diagnosis, for practical reasons we need to be able to name a specific point in time at which a person was considered to have pre-diabetes (or diabetes, or any other health state).

One option is to use the date of the last test as a diagnosis point for both pre-diabetes and diabetes (in other words, the date on which the person fully met the criteria for pre-diabetes / diabetes for the first time). However, the VDR algorithm uses the first date (test visit, hospitalisation event end, pharmaceutical dispensing, or outpatient visit) not the last date, as a proxy diagnosis date.

Initial trials with proxy diagnosis date uncovered problems with using the date of the last lab test as a proxy diagnosis date. If the last date was used, when a cohort of people with pre-diabetes was selected and followed over time to see how many developed diabetes each year, there was a bump of people with diabetes in that first year. Some of these people are the ones where they had two or three tests towards the end of the two year period (and didn't meet the criteria for diabetes) but had one or two tests early in that next two year period and then did meet the criteria for diabetes. It looked like a large proportion of people developed diabetes in the first year following diagnosis but actually, they were just in the first stages of their diabetes diagnosis.

If the first test in a string of lab tests (HbA1c tests for pre-diabetes, HbA1c tests plus two ACR tests for diabetes) is used, the first date is likely to be the same date for diabetes and pre-diabetes – both algorithms go back to (more or less) the same point (date of first test). Using the first test date for the pre-diabetes algorithm definition is consistent with the methods used in the VDR algorithm, and in a practical sense easier to deal with programmatically if both disease algorithms go back to the start of a lab test sequence.

How will we know it works?

When investigating whether or not an algorithm "works", a key question is: what are we expecting? More specifically: Who does the algorithm pick up? What are the patterns by age, sex and ethnic group? To answer these questions, we need another source of information, ideally for the same population. One source of information about the pre-diabetics population in New Zealand is from the Adult Nutrition Survey (ANS) 2008/09 [16], which estimated the prevalence of pre-diabetes in New Zealand by age, sex, and ethnic group and is based the results of blood tests. We will compare the prevalence we calculate using the pre-diabetes algorithm with the ANS pre-diabetes prevalence to gauge how well the pre-diabetes algorithm is working.

We will also use the algorithm to generate pre-diabetes prevalence and incidence rates over time, so that we can examine the potential effects of both the bulking funding of laboratory tests (which increased in recent years) and the introduction of the *More Heart and Diabetes Checks* health target in 2012. We will also investigate rates of progression to type 2 diabetes amongst the pre-diabetic cohort, and compare these with rates from other sources, as another measure of how well the algorithm is working at capturing the pre-diabetic population.

In addition, we can use some statistical techniques (capture-recapture) to generate a pre-diabetes population estimate based on the overlap between the two data sources fro the pre-diabetes algorithm (NMDS and lab tests). We can then compare the capture-recapture estimate with the pre-diabetes population figures derived from the algorithm directly.

What did we find: which algorithm worked best?

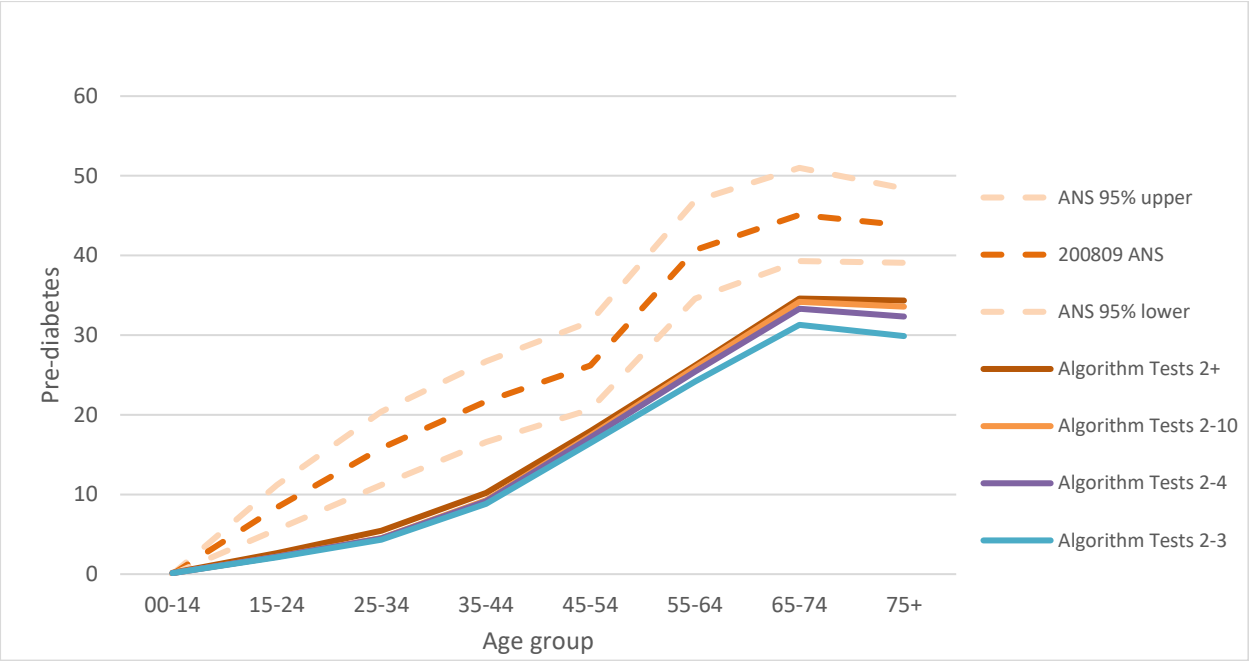
Prevalence

Four different versions of a potential pre-diabetes algorithm were created, and point prevalence of pre-diabetes amongst New Zealand resident health care users was calculated. All four indicators use NMDS diagnosis data E09 in a particular financial year as an indication of pre-diabetes, exclude people with diabetes developing in or before the financial year of interest, and exclude tests occuring during pregnancy (based on NMDS birth event dates and the nine months before this event). The differences between the four algorithm versions relate to the number of HbA1c tests the person has in the period:

- 2 to 3 tests in two years (the financial year noted and the year prior)
- 2 to 4 tests in two years
- 2 to 10 tests in two years
- 2 or more tests in two years

The Figure 4 (below) shows the prevalence of pre-diabetes amongst health service users for these four different variations of the pre-diabetes indicator, compared with the pre-diabetes prevalence estimated from the Adult Nutrition Survey (ANS) 2008/09 [16], which is based the results of blood tests. The data for the pre-diabetes indicator is a point prevalence estimated as at 30 June 2015. Ideally the algorithm prevalences would be extracted for a time period closer to the ANS prevalence period. However, given what we know about HbA1c testing and reporting over time (such as the introduction of current clinical guidelines around screening in 2011, and the More Heart and Diabetes Checks target in 2012), patterns of HbA1c testing in later years may be more likely to be aligned to pre-diabetes status than in earlier years (when the HbA1C test appeared to be less common) .

Figure 4: Esimates of pre-diabetes prevalence by age group, pre-diabetes algorithm and Adult Nutrition Survey 2008/09



The prevalence for the algorithm with 2 or more tests get closer to the ANS figures, but only slightly more that the prevalence for the 2 to 10 test algorithm. To examine the differences between the two algorithms, we drilled down to examine the range of number of tests per person. The data on the number of HbA1c tests for the 2014/15 year for the algorithm where people with 2 or more HbA1c test is shown in Table 2 below:

Table 2: Number of HbA1c tests for people with pre-diabetes identified for 2014/15 by the algorithm

Number of HbA1c Tests	2014/15 people	% Total	% Cumulative
0	215	0.0%	0.0%
1	655	0.1%	0.2%
2	414045	72.5%	72.7%
3	114313	20.0%	92.7%
4	29230	5.1%	97.8%
5	7939	1.4%	99.2%
6	2454	0.4%	99.7%
7	939	0.2%	99.8%
8	493	0.1%	99.9%
9	202	0.0%	99.9%
10	98	0.0%	100.0%
11 to 20	207	0.0%	100.0%
21 to 30	24	0.0%	100.0%
31 to 60	2	0.0%	100.0%
Total	570816	100.0%	100.0%

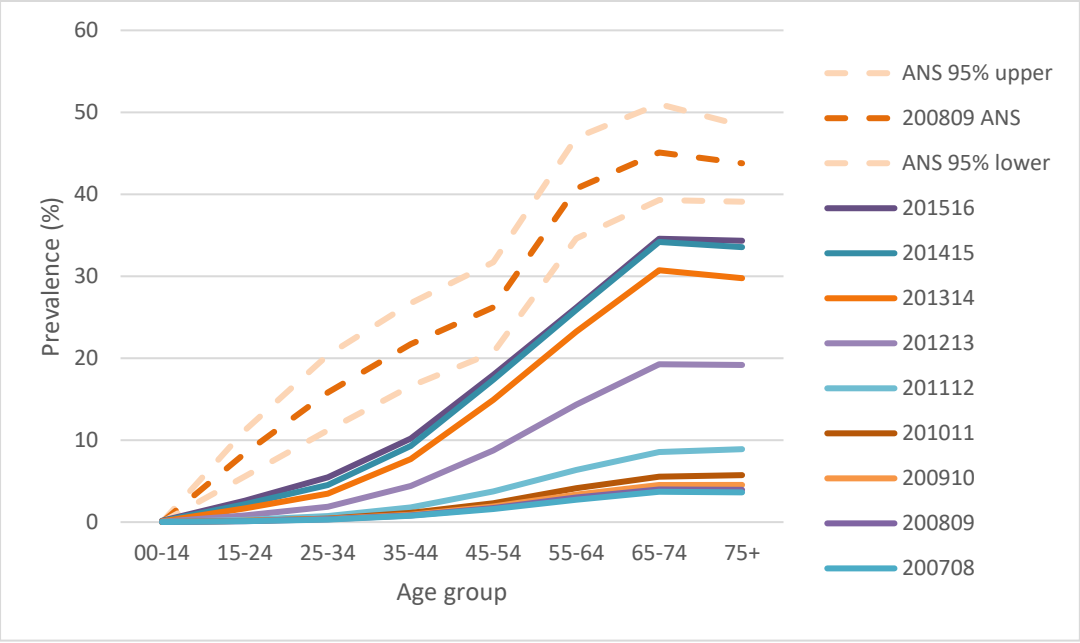
Note that the people with 0 or 1 tests would have been picked up through the NMDS diagnosis data (not the Labs collection part of the algorithm). Other years yeilded very similar patterns. We decided against using the "2 or more" algorithm because there were some people having quite a lot more than ten tests, which may indicate poor data quality or a health condition other than pre-diabetes (but, either way, these people have a testing pattern outside of the NZSSD screening guidelines).

As most people fitted into the ten or fewer range we decided to use the "2 to 10" version of the algorithm: 100.0% of the total number of people detected by the "2 or more" algorithm had had 10 tests or fewer, with only a proportionally very small number having more than 10 tests. In addition, the range of tests for the "2 to 10" algorithm (both upper and lower limits) is linked to guidelines / properties of the test itself: with the minimum level according to guidelines being 2 tests and the upper limit of ten tests being the average amount of time the test actually covers (ten weeks). Further, the prevalence generated from the "2 to 10" algorithm was one of the prevalence estimates closest to the ANS estimates (there wasn't much difference between the 2 to 10 test algorithm estimate and the 2 or more algorithm estimate).

Prevalence over time

Using the "2 to 10" algorithm, we extracted pre-diabetes prevalence over time. The following figure shows the pre-diabetes prevalence by financial year, alongside the prevence estimates based on the Adult Nutrition Survey 2008/09 [16]. The financial year shown is the year in which a pre-diabetes diagnosis was recorded in the NMDS (based on event end date) or the second year in the two-year window examined for the HbA1c testing patterns. If a person developed diabetes (based on the VDR algorithm) before the end of the financial year, that person wasn't included in the pre-diabetes count.

Figure 5: Estimates of pre-diabetes prevalence (% of NZ resident health service users) using the 2-10 algorithm, over time.

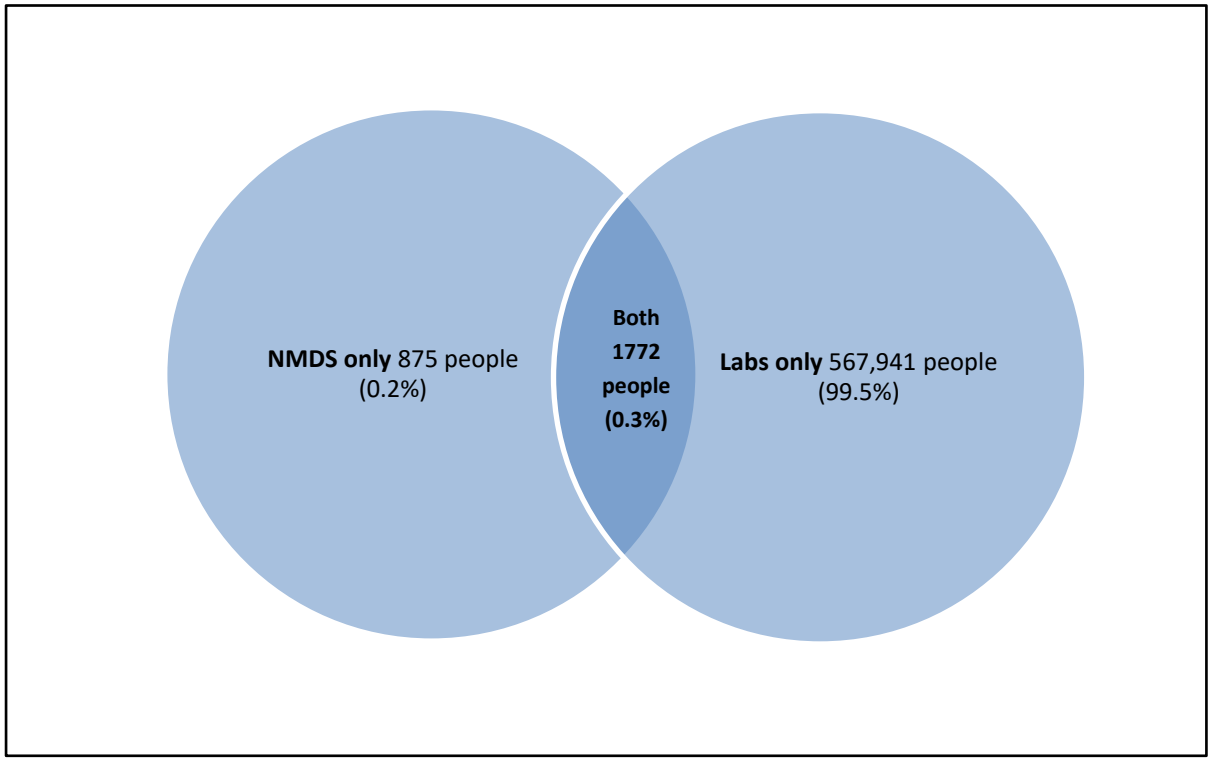


The similar figures for 2013/14, 2014/15 and 2015/16 may indicate that these years are more stable. The increase in prevalence from 2011/12 to 2012/13 levels is likely to reflect the general increases in HbA1c testing around the time the "More Health and Diabetes Tests" target was introduced in 2012, rather than reflecting actual increases in pre-diabetes prevalence.

Overlap between labs and NMDS diagnosis data

As shown in Figure 6 (below) the vast majority of people picked up by the algorithm are coming through from lab testing (99.5%) alone, not from the NMDS diagnosis information.

Figure 6: Pre-diabetes indicator sources for 2014/15 prevalence (2-10 tests algorithm), for health service users with an indication of pre-diabetes, all ages.



Capture-recapture estimate of missing proportion of pre-diabetics

We used a capture-recapture analysis to estimate the proportion of people with pre-diabetes picked up by the pre-diabetes algorithm. Capture-recapture methods have been used in ecological research to estimate the size of animal populations in the wild. These methods are also employed in epidemiology to estimate the size of populations of people with particular health conditions, sometimes with a view to determine the proportion of people missing in a study population or from a disease register (for example, determining the proportion of people with intellectual disability missing from national disability collections [19]).

Table 3: Estimated prevalence of pre-diabetes based on the "2-10 test" algorithm and capture-recapture (CRC) calculations

FinYr	Number of people with pre-diabetes - Algorithm	Number of people with pre-diabetes - CRC Estimate	Population aged 15+ years	Percent capture	Total (15+ years) prevalence - Algorithm (%)	Total (15+ years) prevalence - CRC estimate (%)	Lower 95% CI - CRC estimate (%)	Upper 95% CI - CRC estimate (%)
201112	124077	349356.8	3522345	35.5	3.5	9.9	8.6	11.2
201213	289993	543064.3	3559258	53.4	8.1	15.3	14.5	16.0
201314	486394	750507.7	3585709	64.8	13.6	20.9	20.3	21.6
201415	569341	846616.1	3624639	67.2	15.7	23.4	22.7	24.0
201516	604110	910223.0	3670641	66.4	16.5	24.8	24.2	25.4

The capture-recapture analysis provides a rough guide to the proportion of people with pre-diabetes picked up by the 2-10 test algorithm: around 65-67 percent of people with pre-diabetes are picked up by the algorithm, for the most recent three financial years. The effect of the introduction of the More Heart and Diabetes Checks target in 2012 can be seen by the jump in increases in both estimates of prevalence, as more people had the HbA1c test after the target was introduced.

The estimates of prevalence for 2014/15 were 15.7 percent of the population aged 15 years and older based on the algorithm and 23.4 percent based on the capture-recapture analysis (22.7-24.0 95% confidence interval). The capture-recapture analysis pre-diabetes prevalence gets close to estimates from other sources such as 25.5 percent of the population aged 15 years and over based on the Adult Nutrition Survey [(16)].

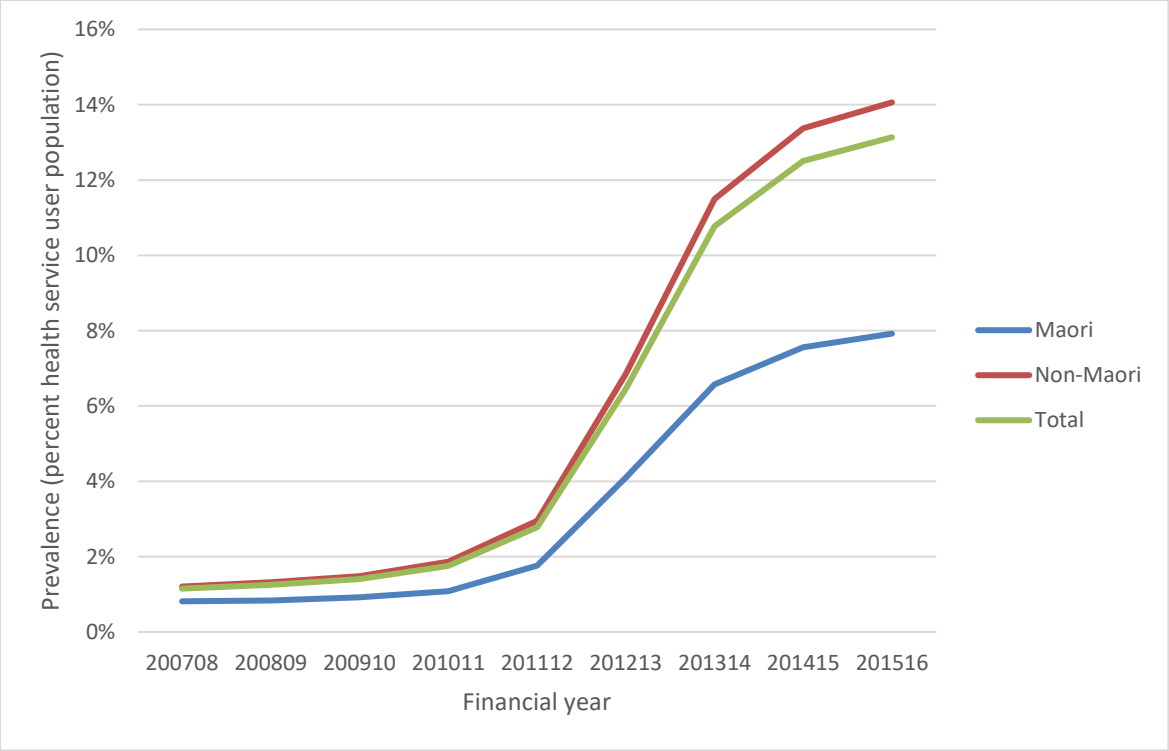
However, the capture-recapture results need to be interpreted with caution as the method relies on a number of assumptions – not all of which are met in this instance. Two key assumptions are met (more or less): the population is closed (people are not coming and going from the population in large numbers); we can reliably link between the two collection points (labs and NMDS hospitalisations – using the encrypted form of the NHI identifier) so people who are picked up with one collection are detected as the same person if picked up in the second collection. However, assumptions that may be only partially met include: the data sources / collection points are not completely independent (for example, someone who had been in hospital may well be more likely to be having lab tests once they are out again), and the people missing from both collections are assumed to be similar to those picked up in the collections (lab tests and hospitalisations) – which may not be the case. The missing people not picked up by the algorithm may not be experiencing symptoms of pre-diabetes and are otherwise healthy – so are not coming into contact with the health system and being tested for pre-diabetes. Conversely, people with pre-diabetes missed by the algorithm may be sicker than those picked up by the algorithm, as their pre-diabetes is not being monitored.

Thinking more generally about who the algorithm misses: we can see in some of the figures above (Figure 4 and Figure 5) that prevalence derived using the algorithm sits at a similar point behind the ANS estimates of prevalence across each the age groups. However, the story is different when we split the data out by ethnicity.

Prevalence by ethnicity

The gap in pre-diabetes prevalence (using the 2-10 HbA1c testing + E09 diagnosis algorithm) between Māori and non-Māori increases over time.

Figure 7: Pre-diabetes prevalence (2-10 test algorithm), by ethnicity and financial year, % of health service user population



This is likely to be tied to the health target that was introduced in 2012. As noted in the evaluation report, inequity of coverage of the Checks health target increased over time following the introduction of the target. Specifically:

"The data shows the coverage gap to have been comparatively small (0.7 percent) when the *Checks* health target commenced in 2012, then expanding, plateauing, and more recently beginning to reduce. One possible explanation for the comparatively small gap at the beginning of the *Checks* health target is that the predecessor health target placed considerable emphasis on diabetes, with 'significant outreach' to Māori populations noted by the Ministry…..The data show that from about mid-2014, a plateau was apparent, suggesting that without conscious, ethnicity-specific, culturally aligned efforts, future reductions in the equity gap associated with ethnicity may be negligible." (Allen & Clarke report, p 30 [14].

Based on data from the Adult Nutrition Survey 2008/09, Coppell et al. [16] calculated pre-diabetes prevalence at 30.4 percent (25.8-35.0, 95%CI) for Māori, 29.8 for Pacific peoples (24.8-34.7, 95%CI) and 24.6 (22.4-26.9, 95%CI) for New Zealand European. The values in the above figure estimate prevalence using this algorithm at around 8 percent for Māori (for 2015/16 data) and 14 percent for non-Māori. Clearly, the non-Māori figure is closer to the ANS estimates than the figure for Māori. We need to remember this gap when extracting rates of others disease amongst pre-diabetics.

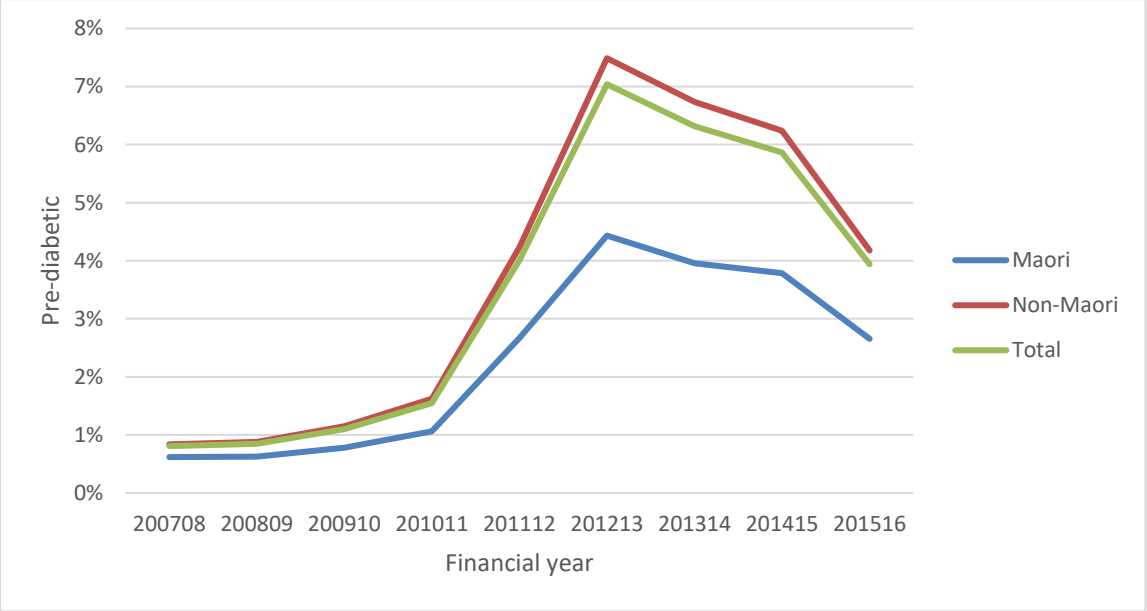
Incidence

Examining incidence over time for an algorithm based on linked health data can sometimes provde clues as to how well an algorithm may be picking up people with the disease of interest. Lots of new cases over time (more than would be expected as the true trend in disease incidence) can indicate that the algorithm didn't work well in the past in terms of coverage (and is picking up prevalent cases as "new" or incident cases – the prevalent pool effect) – and plateaus after a spike in incidence can (but not always) indicate more stability for that period for the algorthim.

Incidence of pre-diabetes was calculated based on the start of a pre-diabetes "episode". Consecutive financial years were linked together where a person met the criteria for the pre-diabetes algorithm (that is, they had 2 to 10 HbA1c tests in two years, and / or an E09 diagnosis for a publicly funded hospital discharge) to form an "episode", treating pre-diabetes as a state that someone can move in and out of. If the person met the algorithm criteria in one financial year and then again a couple of years later (but not in the year in between), these would count as two separate episodes.

The gap between Māori and non-Māori is evident in pre-diabetes incidence, with the introduction of the health target having a similar effect to that seen in the prevalence figures by ethnicity. However, unlike the prevalence gap, the incidence gap is reducing as at the 2015/16 financial year.

Figure 8: Incidence of pre-diabetic episodes (2-10 test algorithm), by ethnicity and financial year



Incidence of diabetes amongst cohorts of pre-diabetics

Another way to examine how well an algorithm is functioning is to investigate the development of rates of disease amongst people identified by the algorithm and compare this with known rates from literature. To test the pre-diabetes algorithm, we examined the incidence of type 2 diabetes amongst a few cohorts of people with indications of pre-diabetes (based in the 2-10 HbA1c test, E09 diagnosis verion of the algorithm).

Teng et al. [17] observed the following cumulative rates of type 2 diabetes incidence amongst a group of people with pre-diabetes in New Zealand: "4.95% (95%CI 4.53–5.42) at three years. This figure was 1.04% (0.88–1.23) at one year, 2.99% (2.69–3.32) at two years, 7.65% (7.00–8.35) at four years and 9.45% (8.52–10.48) at five years." (p.121). These equate to the following year-by-year (non-cumulative) incidence rates:

Table 4: Cumulative and year-by-year type 2 diabetes incidence rates amongst people with pre-diabetes

	year	1	2	3	4	5
Cumulative incidence rates (Teng et al.)	**Incidence (%)**	1.0	3.0	5.0	7.7	9.5
	CI low (%)	0.9	2.7	4.5	7.0	8.5
	CI high (%)	1.2	3.3	5.4	8.4	10.5
Year-by-year incidence rates (calculated based on Teng et al. figures)	**Incidence (%)**	1.0	2.0	2.0	2.7	1.8
	CI low (%)	0.9	1.8	1.8	2.5	1.5
	CI high (%)	1.2	2.1	2.1	2.9	2.1

We extracted similar year-by-year type 2 diabetes incidence rates for a few cohorts of people with pre-diabetes identified by the algorithm. Type 2 diabetes was determined using the same algorithm used in the VDR (i.e. based on data from the national health collections), with some modifications to split out type 1 and type 2 diabetes separately (the VDR includes both types together – see Appendix C for the specific method). On the following page are the incident cases (numbers and percentages) of type 2 diabetes amongst annual cohorts of people with pre-diabetes, over time. The pre-diabetes algorithm used here is the 2-10 HbA1c tests in two years and / or an NMDS diagnosis of E09.

Table 5: Number of incident cases of type 2 diabetes in each financial year, amongst each cohort of pre-diabetics

Cohort year	Population	Pre-diabetics	201011	201112	201213	201314	201415	201516	201617
200910	4377092	61472	3805	3400	2524	1841	1423	725	388
201011	4427091	77458		4271	3405	2478	1922	925	525
201112	4462769	124251			4855	3852	3055	1503	805
201213	4498345	290456				6757	5889	2908	1666
201314	4523594	487319					8164	4326	2499
201415	4561555	570588						5102	2992
201516	4611035	605686							3403

Table 6: Year-by-year incidence of type 2 diabetes, amongst each cohort of pre-diabetics (as percentages; incident cases per year / pre-diabetics in cohort year * 100)

Cohort year	Population	Pre-diabetics	201011	201112	201213	201314	201415	201516	201617
200910	4377092	61472	6%	6%	4%	3%	2%	1%	1%
201011	4427091	77458		6%	4%	3%	2%	1%	1%
201112	4462769	124251			4%	3%	2%	1%	1%
201213	4498345	290456				2%	2%	1%	1%
201314	4523594	487319					2%	1%	1%
201415	4561555	570588						1%	1%
201516	4611035	605686							1%

Because of changes in volumes of HbA1c lab tests associated with the introduction of the More Heart and Diabetes Checks target in 2012, cohorts prior to 2012 may not be as reliable as later years. Post 2012, the incidence rates seem to be steadier, with similar incidence rates each year. The year-by-year type 2 diabetes incidence rates derived from the Teng et al. study [17] was roughly 1-2 percent of people with pre-diabetes each year. Amongst the pre-diabetes cohorts defined using the algorithm, the incidence rates of type 2 diabetes were most similar to the figures derived from Teng et al findings for the final three cohort years (2013-14 to 2015-16), providing further evidence that the pre-diabetes algorithm works best for these years.

Conclusions:

Which algorithm to use to define the pre-diabetes cohort?

The pre-diabetes algorithm to use for BODE^3 work is any New Zealand healthcare user who:

- Has a publicly funded hospital discharge diagnosis E09 (any diagnosis – principal or other), as reported to the NMDS, in the financial year of interest, or
- Has 2 to 10 HbA1c tests in a two year period, as reported in the Laboratory Claims collection, excluding those performed during pregnancy (the window between an NMDS birth event and nine months before the event).

The algorithm excludes people with diabetes (defined using the VDR criteria) diagnosed during or before the financial year of interest. This algorithm was chosen as it:

- Was one of the algorithms which yeilded prevalence figures close to other known pre-diabetes prevalence values (ANS prevalence from 2008/09), with similar pattern by age group ,
- Had upper and lower test limits which aligned with recommended practice for HbA1c testing frequency, and
- Out of all the people with two or more HbA1c tests (and no diabetes diagnosis), most people had 2-10 tests (not many more people were added by including those with more than 10 tests in a two year period).

Caveats for the use of this pre-diabetes algorithm

1. It is worth remembering that the algorithm picks up a group of people who are likely to be those with are “known” or “treated” pre-diabetes. The algorithm will not pick up all people with pre-diabetes. It will not pick up people not interacting with the health system. The missing people may be healthier (asymptomatic, and / or not in contact with health system for other health problems) or they may be less healthy (as their pre-diabetes is not being monitored or treated) than the people picked up by the pre-diabetes algorithm.

2. The pre-diabetes algorithm is recommended for use for the periods 2013/14, 2014/15, and 2015/16; the algorithm is not recommended for periods prior to 2013/14 . The reasons for this are:
    - Current HbA1c testing guidelines (on which the algorithm is based) were released in late 2011.
    - Prior to the introduction of the *More Heart and Diabetes Checks* health target in 2012, HbA1c testing rates were much lower, and probably not in line with current testing guidelines.
    - Prevalence calculated using the algorithm cohort is most similar to other known sources (ANS 2008/09) for these periods,
    - Capture-recapture analysis indicates that algorithm capture in the years 2013/14 to 2015/16 is greater (roughly 65-67 percent of the CRC estimated population) than in earlier periods.
    - Progression to type 2 diabetes for the algorithm cohorts in more recent years is more in line with evidence from other sources

3. The algorithm "works better" for non-Māori, than it does for Māori. Māori prevalence and incidence rates generated using the algorithm are much lower than expected (for example, in relation to ANS 2008/09 prevalence). This lines up with findings from the health target evaluation: the gap between Māori and non-Māori rates of HbA1c testing widened over the period of the health target evaluation, with Māori rates much lower than non-Māori. This means the pre-diabetes prevalence and incidence figures for Māori presented here are not representative of true ethnic differences in pre-diabetes prevalence and incidence.
While the algorithm may capture a smaller proportion of Māori with pre-diabetes, it is still possible to use the algorithm for a denominator in disease rate calculations (i.e. for this work) – as long as this difference is considered when examining disease rates by ethnicity. It is also worth noting that rates for Māori are likely to be less stable because numerators and denominators start to get very small – collapsing demographic categories (for example, combining age groups into wider age bands) may be necessary.

Risk factor distributions

Risk factors included in the pre-diabetes model were those included in the BODE³ DIET Multistate Lifetable Model, where feasible (see 'Technical report for BODE³ DIET intervention and Multistate Lifetable Models' for criteria for risk factor inclusion) and some additional risk factors particularly relevant to those with pre-diabetes.

BMI, blood pressure and total blood cholesterol and dietary risk factors were obtained from the 2008/09 New Zealand Adult Nutrition Survey (2008/09 ANS). As discussed in previous sections, the NZANS took a blood sample for the measurement of glycated haemoglobin (HbA1c). Using the HbA1c measurements the NZANS data was split into those with pre-diabetes, those with diabetes and those without either. The data for those with pre-diabetes was used to populate the pre-diabetes multi-state life-table model. For detailed methods on food consumption data and nutrient intake data from the NZANS please see the 'Technical report for BODE³ DIET intervention and Multistate Lifetable Models'. Risk factor data was weighted within the pre-diabetic population using weightings provided by the NZANS.

Risk factors were split into the same categories that were used in the BODE³ DIET MSLT model (covering the whole NZ population) and the proportion of the pre-diabetic population that fell into each of these categories was calculated for as many of the risk factors in the original DIET MSLT model as was feasible. Midpoints within these categories remained the same as in the original DIET MSLT model. Pre-diabetic specific proportions were calculated for the following risk factors from the DIET MSLT model: BMI; sodium; polyunsaturated (PUFA) fat; fibre; fruit; vegetables; red meat and processed meat intake.

Two risk factors in the original DIET MSLT model could not be calculated:

- Sugar sweetened beverages (intake would need to be adjusted for usual intake as in the original DIET MSLT model due to the high number of people consuming no sugar sweetened beverages in the pre-diabetic population)
- Nuts and seeds (not included due to the high number of people consuming no nuts and seeds in the pre-diabetic population)

Category proportions and midpoints for additional risk factors that were relevant to diabetes prevention interventions were also calculated:

- Systolic blood pressure
- Total blood cholesterol

For many of these risk factors numbers were too small to break the data down by age, sex and ethnicity for these calculations. Data was compared by age, sex and ethnicity and data was broken down by the variables that provided the most variation. For example, if total blood cholesterol varied more by age than by sex then men and women were combined and data was broken down by age to calculate proportions used to populate the model.

Table 7: Risk factors in the Pre-diabetes MSLT model, and which diet disease associations are modelled

	BMI	Systolic blood pressure	Total blood cholesterol	Fruit	Vegetables	PUFA (%TE)	Sodium	Red meat	Processed meat	Fibre*
CHD	√	√	√	√	√	√	√	√	√	√
Stroke	√	√		√	√		√			
Type 2 diabetes	√								√	
Osteoarthritis	√									
Oesophageal cancer	√			√						
Colorectal cancer	√							√	√	√
Breast cancer	√									
Ovarian cancer	√									
Stomach cancer							√			
Lung cancer				√						
Head & neck cancer				√						
Pancreatic cancer	√									
Gallbladder cancer	√									
Thyroid cancer	√									
Liver cancer	√									
Kidney cancer	√									
Endometrial cancer	√									

*Fibre is not modelled at the same time as sources of fibre (fruit, vegetables, nuts and seeds)

Disease rates amongst people with pre-diabetes

Evidence from literature (see below) indicated disease rates amongst people with pre-diabetes differed from the rates in the whole population for the following health conditions:

- coronary heart disease
- stroke
- pancreatic cancer
- endometrial cancer
- stomach cancer

Coronary Heart Disease (CHD)

From the available evidence there appears to be an association between pre-diabetes and CHD. Three studies provide evidence of an association between higher HbA1c and CHD incidence. A 2010 meta-analysis by Sarwar et al [(20)] found a higher risk of CHD incidence with higher HbA1c in non-diabetics (relative risks (RR): 1.20 (95% CI: 1.10–1.31) per 1% higher HbA1c). A 2016 meta-analysis by Huang et al[(21)] found Increases in HBA1c to 39-47 mmol/mol or 42-47 mmol/mol were both associated with an increased risk of composite cardiovascular disease (RR: 1.21 and 1.25, respectively) and coronary heart disease (RR: 1.15 and 1.28, respectively). Yeung et al[(22)] used mendelian randomisation and found that HbA1c was associated with increased coronary artery disease (CAD) risk (Odds Ratio (OR): 1.50 per %, 95% CI: 1.08–2.11).

Two meta-analyses investigated the association between pre-diabetes and CHD mortality. The first found higher risk of mortality with elevated HbA1c in non-diabetics but not diabetics in people hospitalised with CHD.[(23)] Geng et al[(24)] found that in nondiabetic patients with CAD, a high HbA1c level was associated with a higher rate of long-term death (OR: 1.76, 95% CI: 1.44–2.16), and myocardial infarction (OR: 1.69, 95% CI 1.07–2.67).

Stroke

The evidence for an association between pre-diabetes and stroke is less clear. As stated above, the meta-analysis by Huang et al[(21)] found that increases in HBA1c to 39-47 mmol/mol or 42-47 mmol/mol were both associated with an increased risk of composite cardiovascular disease but were not with an increased risk of stroke. The study by Yeung et al[(22)] who employed mendelian randomisation concluded that the association of HbA1c with stroke and its subtypes was less clear given the low number of cases. A meta-analysis of prospective cohort studies by Lee et al[(25)] found no increased risk of stroke after adjustment for established cardiovascular risk factors (RR: 1.08, 95% CI: 0.94 - 1.23) in studies defining pre-diabetes as fasting glucose 100-125 mg/dL (5.6-6.9 mmol/L). However, when pre-diabetes was defined as fasting glucose 110-125 mg/dL (6.1-6.9 mmol/L) the random effects summary estimate showed an increased risk of stroke after adjustment for established cardiovascular risk factors (RR: 1.21, 95% CI:1.02 - 1.44). They concluded that pre-diabetes may be associated with a higher future risk of stroke, but the relative risks are modest and may reflect underlying confounding.

Pancreatic cancer:

From the limited evidence available there appears to be an association between pre-diabetes and Pancreatic cancer.[26; 27]

Pancreatic cancer risk gradually increased with increasing pre-diagnostic HbA1c levels up to an OR of 2.42 (95% CI: 1.33 - 4.39 highest [≥6.5%, 48 mmol/mol] vs lowest [≤5.4%, 36 mmol/mol] category), even for individuals with HbA1c levels within the non-diabetic range.[26]

There was a strong linear dose-response association between fasting blood glucose concentration and the rate of pancreatic cancer across the range of pre-diabetes and diabetes. Sensitivity analysis excluding blood glucose categories in the range of diabetes showed similar results (pooled rate ratio per 0.56 mmol/L increase in fasting blood glucose was 1.15, 95% CI: 1.05 - 1.27), strengthening the association between pre-diabetes and pancreatic cancer.[27]

Endometrial cancer:

From the two studies available, it would appear that there was an association between HbA1c and endometrial cancer. In the first, the hazard ratios (HR) were 4.05 (95% CI: 1.10–14.88) for the moderate HbA1c category (6%–6.9%) as compared with persons having low HbA1c.[28]

In a study investigating breast, endometrial and ovarian cancers in women in Sweden, a significant association between impaired glucose metabolism and an increased risk of endometrial cancer was found.[29]

Stomach cancer:

A meta-analysis[30] on the association between HbA1c and stomach cancer which was based on two studies[28; 31] found an increased risk of stomach cancer with increasing HbA1c (independent of diabetes). The increasing log-linear trend obtained from these studies was statistically significant (P: 0.0171, heterogeneity: 0.9439, I^2=1%).

The DIET MSLT model contains prevalence, incidence, and mortality rates for these diseases for the whole population; these are the disease rates that have been updated to adapt the model to run for a pre-diabetic population. While the disease rate information from the literature search provided evidence of difference between the pre-diabetic and whole populations with respect to these five diseases, a more complete set of disease rates were needed for model inputs. This set of disease rates needed to have standard age, sex, and ethnicity groupings, and be extracted for similar time periods. For this piece of work, we calculated prevalence, incidence, and mortality rates of the five diseases amongst the cohort of people identified using the pre-diabetes algorithm, based on data from the national health collections.

Health condition definitions and data sources

The ICD-10-AM codes used to group conditions for the pre-diabetes disease rate calculations were the same as the groupings used in the DIET MSLT model. The algorithms used for these conditions were similar between the two model, though not identical as the disease rates in the DIET MSLT model were drawn from a number of different sources (some external to the University of Otago) – though most were based on national data collections. In this piece of work, all disease rates were drawn from the national data collections based on the following extract criteria:

Table 8: Health condition definition criteria and national collection data source

Health condition	ICD-10-AM grouping	National health collection source			
		NMDS – publicly funded hospital events only	**Pharmaceutical Claims**	**Cancer Registry**	**Mortality collection – used for mortality and incidence rates only**
CHD	I20-I25	Either: One or more publicly-funded hospital discharge diagnosis of the specified ICD codes (any diagnosis); Or: One or more of the following publicly-funded hospital discharge procedures: ICD-10-AM: 3530400, 3530500, 3531000, 3531001, 3531002, 3849700, 3849701, 3849702, 3849703, 3849704, 3849705, 3849706, 3849707, 3850000, 3850001, 3850002, 3850003, 3850004, 3850300, 3850301, 3850302, 3850303, 3850304, 3863700, 9020100, 9020101, 9020102, 9020103)	Or: Two or more dispensings of one of the following pharmaceuticals (Glyceryl trinitrate 1577, Isosorbide Dinitrate 2377 , Isosorbide mononitrate 2836, Nicorandil 1272, Perhexiline maleate 1949) in the specified financial year		Or: (Used for mortality and incidence rates) Underlying cause of death of one of the specified ICD-10-AM codes
Stroke	G45-G46, I60-I67	Either: One or more publicly-funded hospital discharge diagnosis of the specified ICD codes (any diagnosis);			Or: (Used for mortality and incidence rates) Underlying cause of death of one of the specified ICD-10-AM codes
Stomach cancer	C16			Registrations with the specified ICD-10-AM codes	(Used for mortality rates only) Underlying cause of death of one of the specified ICD-10-AM codes
Pancreatic cancer	C25			Registrations with the specified ICD-10-AM codes	(Used for mortality rates only) Underlying cause of death of one of the specified ICD-10-AM codes
Endometrial cancer	C54-C55			Registrations with the specified ICD-10-AM codes	(Used for mortality rates only) Underlying cause of death of one of the specified ICD-10-AM codes

The criteria were used slightly differently depending on the measure (prevalence, incidence, or mortality rate), detailed in the following section.

Disease rate calculations and modelling

1. **Prevalence, incidence, and mortality "raw" rates extracted from national collections**

As the first step, prevalence, incidence, and mortality rates were calculated for each health condition as a numerator divided by a denominator. Using the criteria in Table 8 (above), the "raw" disease rates were calculated as follows:

Mortality rates

Numerators - were based only on data from the mortality collection, and defined as counts of people with:

- an underlying cause of death coded to one of the ICD-10-AM code groups in the Table 8 (above) and,
- a date of death in the specified financial year, and
- were flagged as having pre-diabetes (defined using the pre-diabetes algorithm) as at the start of the specified financial year.

Denominator – person time in years based on health service user (see Appendix B) years at risk, for people with pre-diabetes (defined using the pre-diabetes algorithm) as at the start of the specified financial year.

Incidence rates

Numerators – a count of the number of **new** cases of the condition in the specified financial year, where the person met one or more criteria listed in Table 8 (above):

- based on just cancer registrations for the cancers, or
- based any one of the data sources listed in Table 8 (above) for CHD and stroke,

out of people flagged as having pre-diabetes (defined using the pre-diabetes algorithm) as at the start of the specified financial year. To work out whether a case was "new" or not, we looked back in time for any evidence of previous cases of the same disease for each person. A case was termed new for a person if there were:

- CHD – no other cases of CHD in the ten years before the specified financial year
- Stroke - no other cases of stroke in the ten years before the specified financial year
- Cancers – the cancer registration date for the particular cancer was in the specified financial year and didn't fall within a period of earlier cancer of the same kind (spanning from the registration date of the earlier cancer to an artificial "cure date" - 6 years for stomach and endometrial cancers, and 5 years for pancreatic cancer after the registration date)

Denominator - person time in years based on health service user years at risk, for people with pre-diabetes (defined using the pre-diabetes algorithm) as at the start of the specified financial year.

Prevalence

A point prevalence estimate as at the end of the specified financial year (that is, 30 June).

Numerators – A count of the number of people who:

- Met the criteria for the specified health condition in that year or
    - In the ten years prior (CHD, Stroke)
    - had a cancer registration date prior to, and a "cure date" after the specified financial year end (stomach, pancreatic, and endometrial cancers)
- were flagged as having pre-diabetes (defined using the pre-diabetes algorithm) as at the end of the specified financial year
- Were alive as at the end of the specified financial year

Denominator – A count of people in the health service user population who were:

- were flagged as having pre-diabetes (defined using the pre-diabetes algorithm) as at the end of the specified financial year
- Were alive as at the end of the specified financial year

Misclassification bias and pre-diabetes status

Because the incidence and mortality rate measures span a year, we needed to determine which point in time to use to define whether or not someone is in the pre-diabetic group relative to when we look for the existance of other conditions / health states. At first we used the end point of a particular financial year to determine pre-diabetic status and looked the incidence or mortality of the condition in the same financial year. For example, we extracted information for people who were flagged as being pre-diabetic as at 30 June 2014/15 (so, given the pre-diabetes algorithm spans two years, we were looking at labs tests across 2012/13 and 2013/14 and hospitalisation diagnoses from 2014/15) and for these people, we calculated their overal mortality rate in 2014/15. However, the mortality rates for pre-diabetics were quite a bit lower than those of the general population who didn't have pre-diabetes or diabetes ("neither disease") – which was unexpected. When we tried using pre-diabetic status as at the start of the financial year, the mortality rates for the pre-diabetic population were similar to that of the "neither disease" population, which was more in line with expected patterns.

Underlying the differences in the pre-diabetes mortality rates above was differential misclassification bias: people who died in 2014/15 were less likely to be part of the pre-diabetic group when we measured it as at the end of the same year because of the way that the pre-diabetes algorithm works (looking at a number of tests over a two year period for labs or over one year for hospitalisation diagnoses). Some pre-diabetics had been misclassified as having neither disease (pre-diabetes or diabetes), making it seem like the pre-diabetics had lower rates of mortality than they actually did. But the problem was, they just had had less time to meet the pre-diabetes algorithm criteria because they had died before they had amassed enough lab tests.

Misclassification bias is a type of information bias. It occurs when people are grouped wrongly in terms of their exposure or disease status [32]. When misclassification bias affects one group to a

different extent than another, this is called *differential* misclassification bias [33]. Differential errors can "deeply undermine epidemiological studies"[32]. Misclassification bias has been identified as common source of error when using routinely collected health data in general, and particularly in algorithms used to indicate chronic disease status [34].

In order to avoid the differential misclassification bias described above, we need to define pre-diabetic status at the start of the period. The best time periods to use to indicate pre-diabetic status are 2013/14 and 2014/15 (based on how well the pre-diabetes algorithm works). In the batch of linked health data we were using for this analysis, the most recent complete financial year of mortality data is 2014/15. This means that to examine disease specific or general mortality rates amongst pre-diabetics, we should use the pre-diabetic status from 2013/14 year (which is basically the pre-diabetic status as at the end of 2013/14 / start of 2014/15) to look at mortality in 2014/15. It follows that we should look at incidence rates for 2014/15 (using the pre-diabetes status based on data from 2013/14).

The prevalence data in this work is not affected by the same differential misclassification bias issue as incidence and mortality rates. Prevalence in this study is point prevalence – measured on one specific day (30 June) – so pre-diabetic status is defined on the same day as the disease numbers are counted.

Output: raw prevalence, incidence and mortality rates

Having extract raw disease rates, we graphed and examined the raw prevalence, incidence and mortality rates by financial year, age, sex, and ethnic group. For more common conditions, such as CHD, the rates were reasonably stable (not varying to a large extent from one age group to the next, for example). For other conditions – mainly because of smaller numbers – the rates were more variable and less stable. To produce a set of more stable rates, we used regression to generate predicted rates based on these "raw" estimates.

2. Predicted rates estimated using regression

We estimated incidence and mortality rates using Poisson regression (or negative binomial where overdispersion was detected), and prevalence using logistic regression.

For all disease rate estimates, we ran a backwards elimination with forced main effects, starting with the following model:

$$\boldsymbol{y = \alpha + \beta_0 + \beta_1 sex + \beta_2 age + \beta_3 ethnicity + \beta_4 year + \beta_5 wide\_age * ethnicity + \beta_6 wide\_age * sex + \beta_7 ethnicity * sex}$$

Where, for incidence and mortality rate estimates, y is a count of new cases of disease (incidence), or deaths from the disease (mortality) with the log of the pre-diabetic population person time at risk as the offset (i.e. representing the denominator for the rates). For estimates of prevalence, y is a count of total cases of the disease amongst pre-diabetics alive at the end of the financial year divided by a count of the total pre-diabetic population alive at the end of the financial year.

The terms in the model were defined as:

- Sex – categorical (male, female)
- Age – categorical, 5 year bands from 25-29 to 90+ years
- Ethnicity – categorical (Māori, non-Māori)
- Year – continuous, from financial years 2007/08 to 2015/16.
- Wide_age – categorical, wider age bands, two sets:
    - broader: (25-44, 45-54, 55-64, 65-74, 75-84, 85+), and
    - broadest: (25-44, 45-64, 65-84. 85+)

We initially ran a backwards elimination with forced main effects for each disease (and measure – incidence, prevalence, mortality rate), based on the general model described above. We then ran additional models with broader and then broadest age groups in place of the five-year age group term. All analyses were conducted using SAS, with results output to MS Excel for visual exploration of the predicted rates (by age, sex, ethnicity and year).

3. Select best modelled rates (describe selection criteria)

The following criteria / questions were used to assess each model for each disease and measure (incidence, prevalence, mortality rates):

Table 9: Criteria used to select regression model of rates of disease amongst people with pre-diabetes

A. Is the model adequate?
1. Did the model converge?
2. How stable are the predicted rates? Do the predicted rates look very bumpy or "zig-zag-y" : May mean need to re-running with broader age-bands
3. Compare the predicted rates and the observed rates. (Look at the plots for each sex*eth group. Do the two lines look broadly similar for most groups?)
4. Check for overdispersion (The value/DF measure for deviance and Person's chi square should be close to 1 for Poisson, if not then could be overdispersed if much greater than 1 or underdispersed if much less than 1)
5. Other sources How do the predicted rates compare to general patterns seen in other populations / publications?
B. Which model is better?
1. Model terms selected What terms were left in the backwards elimination models? Anything surprising left out?
2. AIC (smaller is better) - rank models 1 (best) to 3 (worst)
3. AICC (smaller is better)
4. BIC (smaller is better)
5. Look at predicted versus observed rates Does one model have predicted rates more in line with observed rates?
6. Other sources Compare rates to other known sources? Is one model more like known data than the other?

The questions in Table 9 (above) assess both how well the model works by itself, and which model is "best" out of all the models run for each disease and measure. Table 10 (below) details which model was selected for each disease and measure.

Table 10: Final models selected for each health condition and measure

	Measure		
Health condition	**Incidence**	**Prevalence**	**Mortality rate**
Coronary Heart Disease	NB, 1	LSW, 1	NB, 1
Stroke	P, 1	LSW, 2	P, 1
Stomach cancer	P, 2	L, 2	P, 3
Pancreatic cancer	P, 2	L, 2	P, 3
Endometrial cancer	P, 2, F	L, 2, F	P, 3, F

Key

Code	Meaning
P	Poisson regression
NB	Negative binomial regression
L	Logistic regression
LSW	Logistic regression, scale Williams (to correct for over-dispersion)
1	Backwards elimination + forced main effects + 5 year age groups
2	Backwards elimination + forced main effects + broader age groups
3	Backwards elimination + forced main effects + broadest age groups
F	Females only (no sex terms in model)

Background mortality rates

Using the pre-diabetes algorithm, the health service user population, and the Mortality Collection, we produced estimates of the overall mortality rate of people with pre-diabetes for 2014/15.

First, we extracted "raw" mortality rates for 2014/15, using a similar method to the disease specific mortality rate calculations. For the overall mortality rates, the numerators were counts of people:

- with a date of death (drawn from the mortality collection) in the 2014/15 financial year, and
- who were flagged as having pre-diabetes (defined using the pre-diabetes algorithm) as at the start of the 2014/15 financial year.

The denominator was person time in years based on health service user (see Appendix B) years at risk, for people with pre-diabetes (defined using the pre-diabetes algorithm) as at the start of the 2014/15 financial year.

As the raw rates were not a steady smooth curve across age, sex, and ethnic groups, we used the "raw" mortality rates to produce predicted mortality rates based on Poisson regression. We used a similar approach to the disease rate modelling, running a backwards elimination Poisson regression, with forced main effects, trialling two age band groupings:

- Five-year bands: 0-4, … 100-104, 105+ years
- Broader groups: 0-24, 25-29, … 95-99, 100+ years

Using similar criteria to the disease rate criteria for model selection, the Poisson regression model using the broader age groups was selected, and predicted rates based on this model were output. The predicted rates were then spread out across single years of age, for each unique combination of sex and ethnic group (four groups: Māori males, Māori females, non-Māori males, and non-Māori females). For each of the four sex and ethnic group combinations, we applied Loess smoothing techniques across the age in single years (using the SAS Software LOESS procedure [35]) to produce a smoothed set of predicted mortality rates for 2014/15 amongst people with pre-diabetes.

DISMOD

Disease parameters for the diseases to be added to the pre-diabetes model were inputted to DISMOD II, separately by sex and ethnicity, to generate a mathematically and 'epidemiologically consistent' set of parameters. For example, if the prevalence estimate was too low given what is known about incidence and case-fatality from the disease (and background 'competing' mortality), DISMOD II outputs values that are epidemiologically / mathematically consistent. The process we followed for the new rates specific to those with pre-diabetes was the same as that for the diseases already in the DIET MSLT model. Detailed documentation, including a full example of the process for lung cancer in one of its appendices, can be found in the 'Technical Report for BODE^3 Diet Intervention and Multistate Lifetable Models'.

The DISMOD output rates (in one year age groups) for incidence, prevalence, case-fatality and remission were then used to populate the pre-diabetes MSLT model for stomach, endometrial and pancreatic cancer. For CHD and stroke only incidence, prevalence, and case-fatality were used (i.e. remission was assumed to be zero as these are usually life-long conditions).

Background morbidity rates

In the DIET MSLT model, background morbidity is represented as prevalent years lived with disability (PYLDs). To include PYLDs in the pre-diabetes version of the DIET MSLT model, we created estimates of PYLD for the pre-diabetic population.

Data sources:

The data sources for the PYLD pre-diabetes estimate include:

- Linked data from the national health collections held by the Ministry of Health, extracted in 2018, including:
    - National Health Index (NHI)
    - Primary Health Organisation (PHO) Enrolment collection
    - Laboratory Claims
    - Pharmaceutical Claims
    - National Minimum Dataset (NMDS) - publicly funded hospitalisations only
    - Cancer Registry
    - Mortality Collection
    - National Non-Admitted Patient Collection
    - General Medical Subsidy (GMS) Claims
- The orginal PYLDs for the NZ total population 2011 from the BODE^3 DIET MSLT model
- Tools for estimating comorbidity: SAS macros for the calculation of comorbidity using either the Charlson Index or the M3 index [36], downloaded from here: https://www.otago.ac.nz/wellington/departments/publichealth/research/otago715916.html

Method

Step 1 : Extract the 2013/14 NZ Resident Health Service user population and flag who has an indication of pre-diabetes, diabetes or neither disease as at 30 June 2014

For the NZ resident healthcare user population for the 2013/14 financial year (see definition in Appendix B), we grouped people as to whether they were:

- Pre-diabetic (using the pre-diabetes algorithm) as at the end of 2013/14
- Diabetic (using the VDR algorithm criteria) as at the end of 2013/14
- Neither disease (people who were grouped as pre-diabetic or diabetic).

Step 2: Extract comorbidity scores for 2013/14 NZ Resident Health Service users and attach to the dataset from step 1

We extracted publicly funded hospitalisation data and cancer data for a 5 year period (event end dates / registration dates from 1 July 2009 to 30 June 2014), and created a person level file with all hospital diagnosis and cancer data. Using the comobidity macro developed by Stanley et al. [36], we produced two kinds of comorbidity scores, for each person: a Charlson Index score and a Multi-morbidity (M3) score. These comorbidity scores were then added to the health service user population dataset from step 1.

Step 3: Summarise the person level dataset by age group, sex and ethnicity

For each age, sex, ethnic and disease (pre-diabetes, diabetes, neither, and in total) group, we produced the following summary measures:

- population (number of people) - i.e. Neither_pop, diabetes_pop, prediabetes_pop, whole_pop
- average Charlson Index score - i.e. Neither_avg_CI, diabetes_avg_CI, prediabetes_avg_CI, whole_pop_avg_CI
- average M3 score - i.e. Neither_avg_M3, diabetes_avg_M3, prediabetes_avg_M3, whole_pop_avg_M3

Note that the population figures can also be used to calculate point prevalence for the condition groups (e.g. pre-diabetes prevalence = pre-diabetic population / whole population)

Step: 4: Add in the whole population pYLD from used in the DIET MSLT model and prepare regression input

For the just the whole population summary data from step 3, we added on the PYLDs for the total population from the DIET MSLT model. Then we prepared the table for regression analysis input by applying a logit transformation to the PYLD variable, like this:

LogitPYLD = (log(whole_pop_pYLD/(1-whole_pop_pYLD))

This created a LogitPYLD figure for each age*sex*ethnicity combination (just like there was a regular PYLD figure for each combination as well). We create additional variables for age, and comorbidity score:

- age1 = age as a continuous variable
- age2 = age1 squared
- M3_2 = the M3 score squared
- CI_2 = the Charlson Index score squared

Step 5: Model the relationship between LogitPYLD and age, sex, ethnicity, and comorbidity measure

We used OLS regression (in SAS software, proc genmod [35]) to test 12 different models, based on

LogitPYLD = age * sex * ethnicity * comorbidity

where the defined the terms in different ways:

age	categorical age, continuous age, continuous age squared
sex	categorical variable
ethnicity	categorical variable
comorbidity	average M3 score, average M3 score squared, average CI score, average CI score squared

The criteria for model selection was based on whether the model coverged, the AIC scores (etc), and a visual comparison of graphs (by age, sex, and ethnic group) for the actual and predicted PYLDs (once the logit PYLDs were transformed back into regular PYLDs). The model selected through this process was:

LogitPYLD = categorical age * sex * ethnicity * average M3 score

Step 6: Use the selected model from step 5 to predict PYLDs for different disease populations, based on their specific comorbidity scores

Next we took the pre-diabetes, diabetes and neither disease summaries from step 3 and swapped their average comorbidity score (one at a time) into the place of the comorbidity score in the regression input table (the one for the whole population), yeilding three new regression inputs

- Predibetes group input
- Diabetes group input
- Neither disease group input

Using the stored regression model we selected from step 5, we produced new predicted logit PYLDs for each of these inputs. The LogitPYLDs were transformed back into regular PYLDs using the following equation:

PredictedPYLD for disease group = exp(predicted logitPYLD)/(1+(exp(predicted logitPYLD)))

We then created a weight to apply to predicted population PYLDs for each age * sex * ethnic group:

weight = whole population PYLD / ((neither prevalence *PredictedPYLD for neither)+(diabetes prevalence*PredictedPYLD for diabetes)+(pre-diabetes prevalence*PredictedPYLD for pre-diabetes))

Next we applied the weight to the predicted condition (diabetes / pre-diabetes / neither group) PYLD (for the specific age * sex * ethnic group) to derive final PYLDs for each age, sex, ethnicity and condition (diabetes / pre-diabetes / neither) group. The final PYLDs for the pre-diabetic group are used in the model.

Diabetes: both a disease and a risk factor

It is assumed in MSLT modelling that the incidence rate for a given disease (e.g. CHD) is independent of other diseases (e.g. the presence of diabetes). However, in reality diabetes is associated with increased rates of CHD and stroke (and some cancers), be it by shared common causes (i.e. confounding) or cause and effect (the concern here). For the BODE^3 DIET MSLT model, interventions that change BMI and thence disease incidence are important. BMI is independently associated with each of CHD and diabetes and change in the BMI distribution combined with the relative risk for the BMI→CHD and BMI→diabetes association to give a PIF results in a change in both disease incidence rates. The change in mortality, morbidity and cost rates that result are then 'added' to the overall mortality, morbidity and cost rates in the main life-table. How this is dealt with in the BODE^3 DIET MSLT model and the pre-diabetes model is explained in detail in the 'Technical Report for BODE^3 Diet Intervention and Multistate Lifetable Models' including a full description of how the model alteration was specified in its Appendix: Parameterisation of 'DM as both a risk factor and disease'.

Parameters kept the same as the DIET MSLT model

Disease specific disability rates

Disability rates (DRs) were calculated by dividing the NZBDS's disease-specific pYLDs (adjusted for other co-morbidities, for the year 2006, projected to 2011) by the DISMOD II estimated prevalent cases for all diseases. To estimate the pYLDs in 2011 we applied the following equation:

Generating DRs by dividing pYLDs by prevalent cases for each 5-year age group, for each disease, for each sex by ethnicity, was often too unstable due to sparse data. We therefore aggregated age groupings to ensure the sum of prevalent cases exceeded 10 (e.g. 0-44 year olds were always combined; for common diseases such as CHD and stroke age groupings were: 0-44, 45-54, 55-64, 65-74, and 85+ years; for rare diseases such as pancreatic cancer in Māori males all age groups were combined).

Disease trends

Disease rates in the model are for 2011 only. Some key parameters are known to have increasing or decreasing trends in recent decades – and are likely to have such trends in the near-future. Thus, we also specified future disease incidence and case-fatality as percentage annual change from 2011 to 2026. We assumed that these trends were the same in the pre-diabetic population as they were in the general population so used the same trends as in the DIET MSLT model. For CHD and stroke, we relied on NZBDS projections for annual changes in incidence and mortality. For cancer trends, we relied on our previous modelling of future cancer incidence.[37] Uncertainty around the incidence, case-fatality and remission disease trends were included in the model for all diseases of 1 percentage point SD about the annual percentage change. For details, see the 'Technical report for BODE^3 DIET intervention and Multistate Lifetable Models'.

Disease health system cost inputs

Disease health system costs are identical to those in the BODE^3 DIET MSLT model; we have five types of health system cost:

- Main life-table:
    A. Annual cost to the New Zealand health system for being alive for a given sex and age, and *not* in the last six months of life and *not* concurrently alive with one of the modelled diseases (i.e. diet-related disease). All members of the cohort are assigned this cost; it is the base cost.
    B. Excess cost to A for being in the last six months of life if dying of a disease *other than* one of the modelled diseases (i.e. dying of a non-diet-related disease).
- Disease process life-tables:
    C. Excess cost to A for being in first year of diagnosis of a diet-related disease.
    D. Excess cost to A for being alive with a diet-related disease, and neither in the first year of diagnosis nor in the last six months of life if dying of that disease.
    E. Excess cost to A for being in the last six months of life if dying of a diet-related disease.

Costs were sourced from the New Zealand Health Tracker database for all diseases except diabetes, which was sourced through the Virtual Diabetes Register (VDR). The specific details and equations of how these costs, developed within BODE^3, are calculated are detailed in the following online Report: "Kvizhinadze G, Nghiem N, Atkinson J, Blakely T. Cost Off-Sets Used in BODE^3 Multistate Lifetable Models Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme (BODE^3). Technical Report: Number 15. Wellington: University of Otago, Wellington, 2016" (at: http://www.otago.ac.nz/wellington/otago619391.pdf).

All costs are in 2011 New Zealand dollars. For these costs, see Appendix F in the 'Technical report for BODE³ DIET intervention and Multistate Lifetable Models'.

Checking process

We started with version 2 of the DIET MSLT model, developed over 2018/19 and checked following BODE^3 protocol of checking. For checking of the pre-diabetes MSLT model expected values were run before and after any change that was made and the subsequent change in QALYs were recorded and assessed. When changes in outcomes were of a different magnitude or direction to what was expected based on the model changes, model changes were investigated in more detail.

The following changes were made and checked:

- Changed the population numbers to reflect the population number for pre-diabetics.
- Switched all other risk factors off and then changed the BMI distribution data to the data for the pre-diabetic population. This check was repeated for intake of fruit, vegetables, sodium, PUFA, fibre, red meat, processed meat.
- Added in SBP risk factor spreadsheet with data for pre-diabetics. Checked the effect for a 10mmHg decrease in SBP (risk factor was not in the original DIET MSLT model so just face value check here).
- Added in total blood cholesterol risk factor spreadsheet with data for pre-diabetics. Checked the effect for a 2mmol/L decrease in total blood cholesterol (risk factor was not in the original DIET MSLT model so just face value check here).
- Added in new background mortality rate that is specific to pre-diabetics.
- Added in new diabetes incidence rates.
- Set diabetes incidence to zero for 0–20-year-olds.
- Add in the new background morbidity rates and pYLD rates.
- Set diabetes starting prevalence to zero for all ages.
- Added in new disease data for the diseases that differed between the whole population and pre-diabetics.

References

1. National Centre for Classification in Health (2019) ICD-10-AM/ACHI/ACS. https://www.sydney.edu.au/health-sciences/ncch/icd-10-am.shtml (accessed 23 November 2020)
2. Rankin J, Best K (2014) Disease registers in England. *Paediatrics and Child Health* **24**, 337-342.
3. Cleghorn C, Blakely T, Nghiem N *et al.* (2017) *Technical report for BODE³ intervention and DIET MSLT models, Version 1. Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme. Technical Report no. 16*. Wellington.
4. Forouzanfar MH, Alexander L, Anderson HR *et al.* (2015) Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* **386**, 2287-2323.
5. Ministry of Health (2019) National collections and surveys. https://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys (accessed 23 November 2020)
6. Ministry of Health (2017) National Health Index. https://www.health.govt.nz/our-work/health-identity/national-health-index (accessed 23 November 2020)
7. New Zealand Society for the Study of Diabetes (NZSSD) (2011) *Summary – Screening for type 2 diabetes*.
8. Ministry of Health (2019) National Minimum Dataset (hospital events). https://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/collections/national-minimum-dataset-hospital-events (accessed 23 November 2020)
9. Ministry of Health (2019) Laboratory Claims Collection. https://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/collections/laboratory-claims-collection (accessed 23 November 2020)
10. CareConnect (2020) TestSafe. http://www.careconnect.co.nz/testsafe/ (accessed 23 November 2020)
11. Braatvedt GD, Cundy T, Crooke M *et al.* (2012) Understanding the new HbA1c units for the diagnosis of Type 2 diabetes. *The New Zealand medical journal*.
12. Ministry of Health (2016) *Pre-diabetes: Risk factor management*. Wellington: Ministry of Health.
13. Ministry of Health (2020) Virtual Diabetes Register (VDR). https://www.health.govt.nz/our-work/diseases-and-conditions/diabetes/about-diabetes/virtual-diabetes-register-vdr (accessed 23 November 2020)
14. Allen + Clarke (2016) *More Heart and Diabetes Checks Evaluation.* Wellington: Ministry of Health.
15. Ministry of Health (2014) *Diabetes in Pregnancy: Quick reference guide for health professionals on the screening, diagnosis and treatment of gestational diabetes in New Zealand.* Wellington: Ministry of Health.
16. Coppell KJ, Mann JI, Williams SM *et al.* (2013) Prevalence of diagnosed and undiagnosed diabetes and prediabetes in New Zealand: findings from the 2008/09 Adult Nutrition Survey. *N Z Med J* **126**, 23-42.
17. Teng A, Blakely T, Scott N *et al.* (2019) What protects against pre-diabetes progressing to diabetes? Observational study of integrated health and social data. *Diabetes Res Clin Pract* **148**, 119-129.
18. Health Navigator (2019) Pre-diabetes. https://www.healthnavigator.org.nz/health-a-z/d/diabetes-pre-diabetes/ (accessed 19 July 2019)
19. Ministry of Health (2011) *Health Indicators for New Zealanders with Intellectual Disability.* Wellington: Ministry of Health.
20. Sarwar N, Aspelund T, Eiriksdottir G *et al.* (2010) Markers of dysglycaemia and risk of coronary heart disease in people without diabetes: Reykjavik prospective study and systematic review. *PLoS medicine* **7**, e1000278.

21. Huang Y, Cai X, Mai W *et al.* (2016) Association between prediabetes and risk of cardiovascular disease and all cause mortality: systematic review and meta-analysis. *Bmj* **355**.
22. Au Yeung SL, Luo S, Schooling CM (2018) The impact of glycated hemoglobin (HbA1c) on cardiovascular disease risk: a Mendelian randomization study using UK Biobank. *Diabetes Care* **41**, 1991-1997.
23. Liu Y, Yang Y-m, Zhu J *et al.* (2011) Prognostic significance of hemoglobin A1c level in patients hospitalized with coronary artery disease. A systematic review and meta-analysis. *Cardiovascular Diabetology* **10**, 1-9.
24. Geng J, Zhang Y, Wang B *et al.* (2017) Glycosylated hemoglobin levels and clinical outcomes in nondiabetic patients with coronary artery disease: A meta-analysis. *Medicine* **96**.
25. Lee M, Saver JL, Hong K-S *et al.* (2012) Effect of pre-diabetes on future risk of stroke: meta-analysis. *Bmj* **344**.
26. Grote V, Rohrmann S, Nieters A *et al.* (2011) Diabetes mellitus, glycated haemoglobin and C-peptide levels in relation to pancreatic cancer risk: a study within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *Diabetologia* **54**, 3037-3046.
27. Liao W-C, Tu Y-K, Wu M-S *et al.* (2015) Blood glucose concentration and risk of pancreatic cancer: systematic review and dose-response meta-analysis. *Bmj* **349**, g7371.
28. Travier N, Jeffreys M, Brewer N *et al.* (2007) Association between glycosylated hemoglobin and cancer risk: a New Zealand linkage study. *Annals of oncology* **18**, 1414-1419.
29. Lambe M, Wigertz A, Garmo H *et al.* (2011) Impaired glucose metabolism and diabetes and the risk of breast, endometrial, and ovarian cancer. *Cancer causes & control* **22**, 1163-1171.
30. De Beer J, Liebenberg L (2014) Does cancer risk increase with HbA 1c, independent of diabetes? *British journal of cancer* **110**, 2361.
31. Ikeda F, Doi Y, Yonemoto K *et al.* (2009) Hyperglycemia increases risk of gastric cancer posed by Helicobacter pylori infection: a population-based cohort study. *Gastroenterology* **136**, 1234-1241.
32. Bhopal RS (2002) *Concepts of Epidemiology: An integrated introduction to the ideas, theories, principles, and methods of epidemiology.* Oxford, UK: Oxford University Press.
33. Holman CDJ (2010) Advanced Analysis of Linked Health Data: Topics and Technologies - Short course Workbook.
34. Benchimol EI, Langan S, Guttmann A *et al.* (2013) Call to RECORD: the need for complete reporting of research using routinely collected health data. *Journal of clinical epidemiology* **66**, 703-705.
35. SAS Institute Inc (2015) The SAS system for Windows. Release 9.4 M3. Cary, NC, USA: SAS Institute Inc.
36. Stanley J, Sarfati D (2017) The new measuring multimorbidity index predicted mortality better than Charlson and Elixhauser indices among the general population. *J Clin Epidemiol* **92**, 99-110.
37. Costilla R, Atkinson J, T B (2011) *Incorporating Ethnic and Deprivation Variation to Cancer Incidence Estimates over 2006-2026 for ABC-CBA*. *Burden of Disease Epidemiology, Equity and Cost-Effectiveness Programme*. Wellington: Department of Public Health, University of Otago.

Appendices

Appendix A – Screening Guidelines from the New Zealand Society for the Study of Diabetes [7]

"What to do following a screening test for type 2 diabetes

Result	Action	Why
Symptomatic		
HbA1c ≥ 50 mmol/mol and, if measured, Fasting glucose ≥7.0 mmol/L Or Random blood glucose ≥11.1mmol/L	No further tests required	Diabetes is confirmed
Asymptomatic		
HbA1c ≥ 50 mmol/mol and, if measured, Fasting glucose ≥7.0 mmol/L Or Random glucose ≥ 11.1 mmol/L	Repeat HbA1c or a fasting plasma glucose	Two results above the diagnostic cut-offs, on separate occasions are required for the diagnosis of diabetes*
HbA1c 41-49 mmol/mol and, if measured, Fasting glucose 6.1–6.9 mmol/L	Advise on diet and lifestyle modification. Repeat the test after 6-12 months	Results indicate 'pre-diabetes' or impaired fasting glucose*
HbA1c ≤ 40 mmol/mol and ,if measured, Fasting glucose ≤6 mmol/L	Retest at intervals as suggested in cardiovascular risk factor guidelines	This result is normal

* When HbA1c and fasting glucose are discordant with regard to diagnosis of diabetes, repeat testing at an interval of 3-6 months is recommended. The test that is above the diagnostic cut point should be repeated – if the second test remains above the diagnostic threshold then diabetes is confirmed. If the second result is discordant with the first then subsequent repeat testing at intervals of 3-6 months is recommended. Patients with discordant results are likely to have test results near the diagnostic threshold."

Appendix B – Health Service User criteria – Population denominator

Data sources

Linked data from the national health collections held by the Ministry of Health, extracted in 2018, including:

- National Health Index (NHI)
- Primary Health Organisation (PHO) Enrolment collection
- Laboratory Claims
- Pharmaceutical Claims
- National Minimum Dataset (NMDS) - publicly funded hospitalisations only
- Cancer Registry
- Mortality Collection
- National Non-Admitted Patient Collection
- General Medical Subsidy (GMS) Claims

Method

People are defined to be part of the NZ resident health service user population for the specified financial year, if they

- have at least 1 health system contact (including PHO enrolment) during the financial year and are flagged as a NZ resident on the NHI, or
- have at least 2 health system contacts, three or more months apart during the financial year and are not listed as NZ resident on the NHI.

Appendix C - Definitions for type 1 and type 2 diabetes, extracted from the national health collections (using the algorithm for the Virtual Diabetes Regiser – VDR as a starting point)

Type 1 diabetes

Anyone who meets the VDR definition of diabetes, and then is coded has type 1 if they meet the following criteria:

1) NNPAC attendance

One or more attendances for services provided under the type I purchase unit codes

M20010 (High risk type I diabetes support),

M20015 (High risk type I diabetes support for up to 18 year olds).

2) NMDS / PHARMS

If the person has discharges (NMDS) for type 1 diagnoses and type 2 diagnoses, and no oral hypoglycaemic / metformin dispensings, and one or more insulin dispensings,

- if they are aged 0-14 years then they are coded as type I
- if they are 15 years or older and 95% or more of their discharge / diagnosis events are for type I (out of type I and II in total)

then they are coded as type I

If the person has no type I or type II discharge events, no oral hypoglycaemic / metformin dispensings, and has one or more insulin dispensings then they are coded as type I

But if someone has one to two type I discharge events and none for type II, no oral hypoglycaemic / metformin dispensings,
and no insulin dispensings then they are not coded as type I (they are specifically excluded because there is not evidence of insulin treatment -possible coding F34).

If, after coding all the above, the type is still unknown and the person has had a dispensing of insulin before their 10th birthday
they are coded to type 1 (regardless of dispensings of metformin / oral hypoglycaemics).

Type 2 diabetes

Based on the VDR algorithm, but then filtered out the Type 1 diabetics, leaving the Type 2 diabetics (i.e. by subtraction)