

# BMJ Open Testicular Cancer in New Zealand (TCNZ) study: protocol for a national case-control study

Jason K Gurney,<sup>1</sup> James Stanley,<sup>1</sup> Katherine McGlynn,<sup>2</sup> Lorenzo Richiardi,<sup>3</sup> Caroline Shaw,<sup>4</sup> Richard Edwards,<sup>4</sup> Tony R Merriman,<sup>5</sup> Bridget Robson,<sup>4</sup> Jonathan Koea,<sup>6</sup> Melissa McLeod,<sup>4</sup> Martin A Kennedy,<sup>7</sup> Diana Sarfati<sup>1</sup>

**To cite:** Gurney JK, Stanley J, McGlynn K, *et al.* Testicular Cancer in New Zealand (TCNZ) study: protocol for a national case-control study. *BMJ Open* 2018;**8**:e025212. doi:10.1136/bmjopen-2018-025212

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-025212>).

Received 6 July 2018  
Revised 10 July 2018  
Accepted 12 July 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Cancer and Chronic Conditions (C3) Research Group, Department of Public Health, University of Otago, Wellington, New Zealand

<sup>2</sup>Division of Cancer Epidemiology & Genetics, National Cancer Institute, Maryland, USA

<sup>3</sup>Cancer Epidemiology Unit, University of Turin, Turin, Italy

<sup>4</sup>Department of Public Health, University of Otago, Wellington, New Zealand

<sup>5</sup>Department of Biochemistry, University of Otago, Dunedin, New Zealand

<sup>6</sup>Department of Surgery, Waitemata District Health Board, Auckland, New Zealand

<sup>7</sup>Department of Pathology, University of Otago, Christchurch, New Zealand

## Correspondence to

Dr Jason K Gurney;  
[jason.gurney@otago.ac.nz](mailto:jason.gurney@otago.ac.nz)

## ABSTRACT

Testicular cancer (TC) is by far the most common cancer to affect young men; however, the exposures that cause this disease are still poorly understood. Our own research has shown that Māori men have the highest rates of this disease in New Zealand—a puzzling observation, since internationally TC is most commonly a disease of men of European ancestry. These trends provide us with a unique opportunity: to learn more about the currently unknown exposures that cause TC, and to explain why Māori have the highest rates of this disease in New Zealand. Using epidemiology and genetics, our experienced research team will conduct a nationwide study which aims to answer these internationally important questions.

**Aim of study** The overall aim of the current national case-control study is to identify the key exposures in the development of TC in New Zealand, and explore which factors might explain the difference in the incidence of TC between Māori and non-Māori.

**Methods and analysis** Outside of our own investigations into cryptorchidism, we still do not know which exposures are driving the significant incidence disparity between ethnic groups in NZ. The aim of the proposed research is to use a population-based case-control study to identify the key exposures in the development of TC in New Zealand. We will recruit 410 TC cases and 410 controls, and collect (1) environmental exposure data, via interview and (2) genetic information, via genome-wide genotyping.

**Ethics and dissemination** Ethical approval for this study was sought and received from the New Zealand Ministry of Health's Health and Disability Ethics Committee (reference # 17/NTA/248). Following a careful data interpretation process, we will disseminate the findings of this study to a wide and varied audience ranging from general academia, community groups and clinical settings, as well as to the participants themselves.

## INTRODUCTION

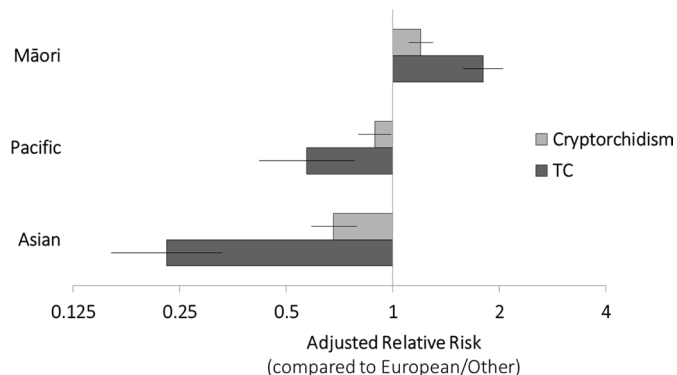
Testicular cancer (TC) is the most common cancer to affect young men, with disease incidence peaking between 15 and 40 years of age.<sup>1</sup> Incidence rates of TC have been increasing rapidly over time, both internationally<sup>2</sup> and in New Zealand.<sup>3</sup> Despite a large body of research focused on this cancer, TC

## Strengths and limitations of this study

- This study is the first to investigate the aetiology of testicular cancer in New Zealand, a country with highly unusual ethnic trends in the incidence of this disease.
- This is a national study, using robust cancer registry data to identify cases of testicular cancer. It will use a case-control design, which is the most efficient means of addressing our research questions.
- One limitation of the study is that, due to the nature of the perinatal and early-childhood exposures we are examining, there is some risk of social stigmatisation and psychological distress of participants (particularly the mothers of cases/controls).
- Linked to this, there is also some risk of non-reporting of exposures that are considered socially undesirable, as well as some risk of differential recall/reporting between cases and controls.
- There are general limitations inherent to most case-control studies, including the potential for low recruitment rates (particularly of controls), and retrospective reporting of exposures which could be differentially reported by cases and controls (recall bias).

remains poorly understood—both in terms of what causes the disease, and also why certain populations are more likely to develop it than others.

Regarding higher risk for particular groups, there are several examples of puzzling epidemiological patterns for TC; for example, despite geographical proximity and very similar population demography, rates of disease among Norwegian men are two and a half times higher than those observed in Finland.<sup>4</sup> In New Zealand, a disproportionately large number of TC cases occur in Māori men, with relative incidence rates 80% higher in this ethnic group compared with European/other men (adjusted relative risk 1.80, 95% CI 1.58 to 2.05).<sup>5</sup> This is in stark contrast to other countries, where white



**Figure 1** Comparison of relative risk (95% CI) of cryptorchidism 1 and testicular cancer (TC) 3 in the New Zealand context, by ethnicity (data from Gurney *et al*<sup>514</sup>).

men—particularly those who can trace their ancestry back to Northern Europe<sup>6</sup>—have by far the highest rates of disease compared with other ethnic groups.<sup>7,8</sup> An additional factor is that Pacific New Zealanders have very low rates of TC compared with Māori and non-Māori/Pacific/Asian men—a rare (if not singular) example where Māori and Pacific peoples do not move in parallel with respect to the incidence of a given disease.<sup>3,5</sup> All of these unusual epidemiological trends remain unexplained.<sup>6–8</sup>

Aside from young age, the risk factors which have been shown to be most strongly associated with TC incidence include prior cancer of the opposite testicle,<sup>9</sup> family history of TC<sup>10</sup> and history of undescended testis also known as cryptorchidism.<sup>10</sup> Given the rarity of these risk factors, they can only possibly explain a small proportion of cases.<sup>11</sup>

### Risk factors for TC

Other than age, cryptorchidism is the risk factor most strongly associated with TC development.<sup>1</sup> It is known that men with a history of cryptorchidism have a three to sixfold increase in the risk of TC compared with those with no history of cryptorchidism.<sup>12–13</sup> Our previous research has shown that ethnic trends in the incidence of cryptorchidism mirror those found for TC (see figure 1), whereby Māori boys are significantly more likely to have this condition than non-Māori/Pacific/Asian boys (rate ratio 1.24, 95% CI 1.15 to 1.34).<sup>14</sup> Our observations of a parallel between TC and cryptorchidism incidence in New Zealand strongly suggest that ethnic differences in rates of cryptorchidism are associated with differing rates of TC—directly via a causal pathway and/or indirectly via common risk factors. Since the factors involved in the aetiology of cryptorchidism occur prenatally,<sup>15</sup> it follows that at least some of the exposures driving these observed disparities occur prenatally.

Since the aetiology of TC remains obscure, a number of other potential risk factors have been investigated which could be important in the development of TC. These risk factors can be broadly grouped as either environmental or genetic, with environmental risk factors subclassified as prenatal and postnatal. Of these risk factors, the

current study is most concerned with those which could plausibly explain the disparity observed between Māori and non-Māori New Zealanders with respect to the incidence of TC and/or cryptorchidism.

### Aims and research questions

The overall aim of the current study is to identify the key exposures in the development of TC in New Zealand, with a focus on those factors which might explain the incidence disparity between Māori and non-Māori men. By identifying these points of difference, we will likely increase international understanding of the key exposures which lead to the development of TC,<sup>3,14</sup> making New Zealand a prime location for this important aetiological research.

The current study will aim to address the following questions:

1. How common are potential environmental risk factors for the development of TC in New Zealand, and how strongly are they associated with TC development in this context?
2. How does the prevalence of these environmental risk factors for TC differ by ethnicity, and to what extent might these risk factors explain the disparity in TC incidence between Māori and non-Māori men?
3. How common are the known genetic risk factors for TC (eg, specific KITLG alleles) among New Zealand men, and how strongly are they associated with TC development in this context?
4. How does the prevalence of these known genetic risk factors differ by ethnicity, and to what extent might these risk factors explain the disparity in TC incidence between Māori and non-Māori men?
5. Is there evidence for additional (as-yet undiscovered) predisposing genetic variants in Māori men, particularly at the known loci? (To be answered in a future study.)

## METHODS AND ANALYSIS

### Study design

In order to answer our research questions, we will conduct a population-based case-control study in New Zealand. We will collect data on exposures occurring prenatally and in the prediagnostic period for cases and an equivalent period for controls using self-report from cases/controls and their mothers. We will also collect genetic data.

### Questionnaire development

In order to develop the study questionnaires, we conducted a literature review of the key biologically plausible exposures in the development of TC. Searches of electronic databases were conducted using Scopus, using relevant keywords and Boolean search terms. We also requested and received copies of study questionnaires from the corresponding authors of several other relevant published studies.

When developing the questionnaires, we focused on purported risk factors for TC where there was either

clear or mixed evidence of an association with TC in the literature, as well as evidence of significant differential exposure between Māori and non-Māori. Both study questionnaires were developed over several years, based on previous questionnaires from other TC case-control studies,<sup>16–18</sup> other questionnaires including the New Zealand Health Survey,<sup>19</sup> and input from experts and key stakeholders.

The case/control questionnaire includes questions on: (1) history of cryptorchidism, inguinal hernia or hypospadias, including whether these conditions were corrected and at what age; (2) active and passive exposure to cannabis; (3) active and passive exposure to tobacco smoke, including known family history of tobacco use and type of cigarette (roll-your-own or factory manufactured); (4) age at puberty and early sexual experiences (as proxies for puberty/sexual maturity); (5) occupational history, including pesticide exposure (among other exposures); (6) fertility, (7) achieved adult height and (8) known family history of TC and/or cryptorchidism. We have also included questions on alcohol intake, primarily with a view to controlling for this exposure as a potential confounder of the relationship between tobacco and cannabis exposure and TC development.

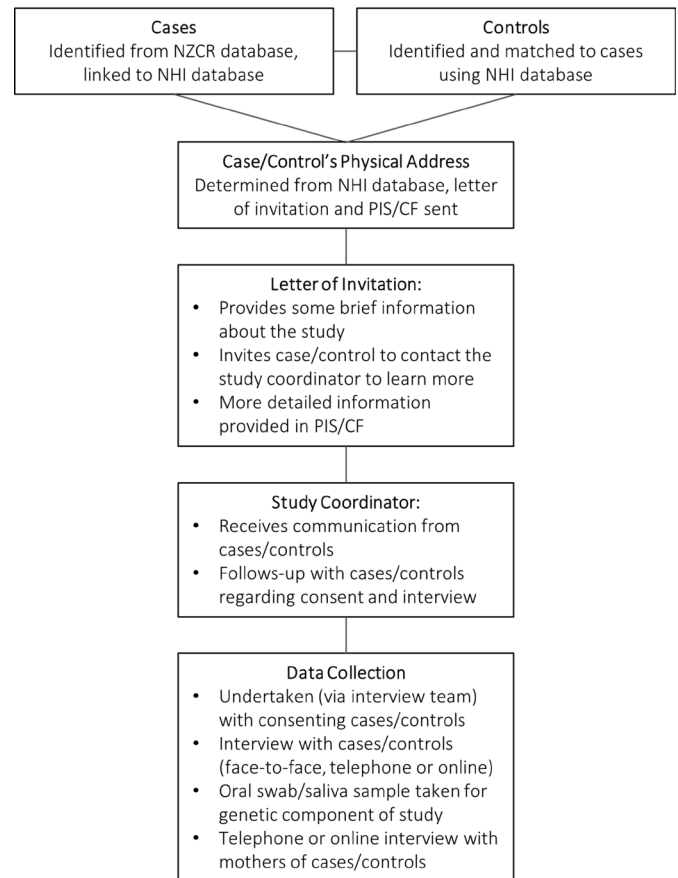
The questionnaire for the birth mother of the case/control will include questions on: (1) maternal (and paternal) smoking, including type of cigarette; (2) gestational age at delivery; (3) birth weight; (4) maternal use of cannabis and other drugs and (5) recollection of any other unusual prenatal events, including nausea beyond the first trimester, excessive maternal weight gain and bleeding. In addition, mothers will be asked for their recollection of (1) cryptorchidism, inguinal hernia or hypospadias occurrence in their son, including whether this condition was corrected and at what age; (2) age of son's puberty and (3) family history of TC and/or cryptorchidism. We have also included questions on alcohol intake during pregnancy, for the reasons given above.

## Participants

### Case identification

Patients diagnosed with TC ('cases') will be identified via unique identifier (National Health Index (NHI) number) using the New Zealand Cancer Registry (NZCR). The NZCR is a national collection of all incident cases of malignancies in New Zealand, with the exclusion of squamous and basal cell malignancies.<sup>20</sup> Only cases aged between 16 and 49 will be included in this study. The NHI number is attributed to all New Zealanders who use healthcare services (approximately 98% of the population), and the NHI database includes additional information including name, address, date of birth, place of birth, gender, residency status, ethnicity and date of death (if relevant).<sup>21</sup> Indigenous Māori and non-Māori TC incident cases diagnosed over the 2010–2016 period will be the base for participant recruitment (approximately 1050 cases or 150 cases/year<sup>22</sup>). Once identified via the NZCR, TC cases will be linked to the NHI database

## NHI controls



**Figure 2** Strategy for identifying and recruiting cases and controls. NHI, National Health Index; NZCR, New Zealand Cancer Registry; PIS/CF, Participant Information Sheet/Consent Form.

via their unique NHI identifier. If the participant has no recorded date of death on the NHI, the process of case recruitment will begin (figure 2). Given the high survivability of TC in New Zealand (5-year survival: European/other 98%, Māori 92%),<sup>5</sup> coupled with the young median age at diagnosis (European/other 38 years, Māori 31 years),<sup>5</sup> the proportion of men who would have died since their diagnosis is expected to be less than 5%.

### Control identification

For every recruited case, one age-matched and ethnicity-matched control will be recruited from the general population. Controls will be identified from the NHI database, and frequency matched to cases on the basis of age (5-year age groups) and ethnicity (Māori/non-Māori).

### Case/control recruitment

Once cases and controls have been linked to the NHI, a letter of invitation and information about the study will be sent to each potential case and control, inviting individuals to contact the study coordinator to discuss the study and obtain consent.



### Interview of cases/controls

Interviews with eligible participants will be available face-to-face interview, by telephone or online. If the participant decides to complete the survey over the telephone or online, a separate face-to-face appointment will be made with them in order to collect the saliva sample required for the genetic component of the study.

### Recruitment of mothers of cases/controls

At the time of the interview with cases/controls, the participants will be asked for their consent for the study team to contact their mothers. If the mother is interested in participating, the study coordinator will make arrangements regarding an interview either by telephone or online. Cases/controls will not be excluded if their mother is unable or unwilling to participate (see [table 1](#) regarding approaches to managing potential bias from missing maternal data).

### Data collection

Two forms of information will be collected for this case-control study: (1) interview information and (2) genetic information. This section further details the information to be gathered during data collection.

Interview data will be used to determine those prenatal and postnatal environmental exposures which are more common among those with TC than those without, as well as ethnic differences in exposure prevalence. Two questionnaires have been developed to identify the prevalence of exposure among cases and controls to a number of TC risk factors. One questionnaire will be answered by the case/control (Case/Control Questionnaire), and a second questionnaire will be answered by the mother of the case/control (Mother of Case/Control Questionnaire). The study questionnaires, including the source of individual questions, where relevant, are available on request from the lead author.

Genetic data will be used to determine the prevalence of risk alleles among New Zealand men, and how the prevalence of risk alleles differs between cases and controls, stratified by ethnicity. DNA samples will be taken from cases/controls via saliva sample, using the Oragene OG-500 kit (DNA Genotek, Ottawa, Canada). Collection of DNA via saliva at the time of the interview removes the need for participants to visit a blood collection centre, and also removes the need for the use of invasive needles. The sample will be taken by interviewers and packaged in laboratory bags, which will have a sticker attached to them that includes only the unique study identifier for the participant (ie, no identifiable information). The laboratory bag will then be delivered to a genetics laboratory for preparation of genomic DNA and subsequent analysis. At this laboratory, genome-wide genotyping will be performed with only those ~40 loci already known to be associated with TC by genome-wide association studies specifically interrogated.<sup>23–26</sup> The genotyping will be performed on an appropriate GWAS platform such as the Illumina CoreExome array. The primary reason

for using a genome-wide platform to collect the data on the ~40 known TC loci is as follows: (1) genome-wide genotyping is a practical cost-effective way to genotype multiple single nucleotide polymorphisms over multiple candidate loci (research questions 3 and 4) and (2) genome-wide genotyping data could, in future, allow a full genome-wide association study which would allow us to examine our data for evidence of additional variants in Māori that might explain our observed disparity (eg, a rare, highly penetrant variant; research question 5). Access to genomic data for this future genome-wide association study would be governed by our Kaitiakitanga (Guardianship) Group (see the Data management section). Following genotyping, the resulting raw data will be processed by an experienced bioinformatician, and analysis will be carried out in conjunction with the wider team.

### Transport, storage and analyses of genetic specimens

We have dedicated substantial time and resources into developing robust protocols around collection and transport of samples, built on the advice and steering of senior Māori health researchers and key stakeholders. We have decided that samples (and/or the resulting genomic data) will not be sent overseas to large analysis facilities (which would be highly cost-effective)—rather, all handling of specimens and analyses will be conducted in New Zealand. A unique study identifier (and no other identifying information) will be used to identify samples during transport and analysis. Importantly, specimens collected for this study will not be used for any purpose other than those declared to participants during the study consent process.

To mitigate the potential risk of losing valuable information by destroying all biospecimens, we will ask participants for permission to retain the genome-wide genotyping data collected for the study beyond the end of the study, access to which will be governed by our Kaitiakitanga (Guardianship) Group. As mentioned above, retaining this genome-wide genotyping data could, in future, allow a full genome-wide association study which would allow us to examine our data for evidence of additional variants in Māori that might explain our observed disparity.

### Data analysis plan

#### Interview data

In addition to descriptive analyses, multivariable logistic regression methods (non-conditional, due to frequency matching approach) will be used to determine the odds of TC among those exposed to a given exposure compared with those who were not exposed, adjusted for confounding factors. This regression analysis will be completed both for the whole study population (ie, all cases/controls), and also stratified by ethnic group (Māori and non-Māori). From these analyses, ORs and their confidence limits will be calculated.

**Table 1** Study risks and approach to their management

Potential risk	Approach to risk management
1. Insufficient total cases/controls.	<ul style="list-style-type: none"> <li>▶ Maximise response rate by creating multiple possible modes of participation.</li> <li>▶ Potentially extend period of case definition (beyond 2010–2016) if necessary.</li> </ul>
2. Insufficient Māori cases/controls, that is, to allow stratified analysis.	<ul style="list-style-type: none"> <li>▶ Maximise response rate by creating multiple possible modes of participation.</li> <li>▶ Potentially extend period of case definition (beyond 2010–2016) if necessary.</li> <li>▶ Increase oversampling of Māori (and thus no of Māori recruited into the study).</li> <li>▶ Potentially extend period of data collection.</li> </ul>
3. Selection bias of cases, that is, testicular cancer (TC) cases included in study are not representative of TC cases generally in terms of exposures.	<ul style="list-style-type: none"> <li>▶ Should be small as TC is uncommon and all diagnosed cases appear on the New Zealand Cancer Registry (NZCR).</li> <li>▶ Characteristics of consenting cases (age, ethnicity, deprivation) can be compared with all NZCR TC cases to see if there are any systematic differences.</li> <li>▶ If clear differences are observed, we will consider and document how this difference may impact the exposure information collected from cases.</li> </ul>
4. Selection bias of controls, that is, controls included in study are not representative of source population in terms of important exposures (eg, those who agree to participate may systematically differ in some way to those who do not).	<ul style="list-style-type: none"> <li>▶ Characteristics of consenting controls (age, ethnicity, deprivation) can be compared with the general population to see if there are any systematic differences.</li> <li>▶ If clear differences are observed, we will consider and document how this difference may impact the exposure information collected from controls.</li> </ul>
5. Information bias—case/control misclassification, for example, cases incorrectly diagnosed as TC, controls incorrectly diagnosed as non-cases.	<ul style="list-style-type: none"> <li>▶ Because cancers registered on the NZCR are histologically confirmed, and because we will ask cases and controls whether they had a previous TC diagnosed prior (or since) the study period (2010–2016), risk of case or control misclassification is very minimal.</li> </ul>
6. Information bias—poor recall of some exposures by cases and controls (non-differential) especially early childhood exposures.	<ul style="list-style-type: none"> <li>▶ Ask cases/controls questions that they are likely to recall, and minimise questions that refer to early to mid-childhood</li> <li>▶ Ask mothers of cases/controls questions that they are likely to recall (eg, smoking during pregnancy or exposure to cannabis), and minimise questions that are difficult to recall (eg, heaviness of bleeding during menstrual cycle prior to index pregnancy).</li> </ul>
7. Information bias—differential recall of those exposures that are asked of both mothers and sons.	<ul style="list-style-type: none"> <li>▶ Give way rule: if exposures pertain to pregnancy and childhood, exposure information from mothers will be preferred to that from sons when interpreting data (although both will be presented).</li> <li>▶ Give way rule: if exposures pertain to adolescence, exposure information from adolescence onwards, exposure information from sons will be preferred to that from mothers.</li> </ul>
8. Information bias—non reporting of exposures that are considered socially undesirable (eg, cannabis use by cases/controls, smoking during pregnancy by mothers).	<ul style="list-style-type: none"> <li>▶ Prior to answering questions about cannabis use/exposure, cases and controls will be informed that their answers will be completely confidential within the study team.</li> <li>▶ Cases/controls: In the first instance, encourage cases/controls to complete the interview face to face. When asking questions for which the answer may be socially undesirable, the interviewer can pass digital tablet to the case/control, so that they can answer the question(s) without the interviewer knowing their answer.</li> <li>▶ Mothers: In the first instance, encourage mothers to complete the study questionnaire online (ie, without an interviewer asking the questions).</li> </ul>
9. Information bias—language difficulties and cultural differences creating non-differential recall of exposures by cases and controls.	<ul style="list-style-type: none"> <li>▶ Using an experienced interviewing service with a track record of cultural competency.</li> <li>▶ Where possible, match interviewers to participants on ethnicity (Māori/non-Māori).</li> </ul>
10. Information bias—differential recall of exposures by cases and controls.	<ul style="list-style-type: none"> <li>▶ Use identical questions for cases and controls.</li> <li>▶ Do not provide information on specific hypotheses being tested.</li> <li>▶ Do not ask questions about exposures immediately prior to illness (for cases), since this is a different period to controls.</li> </ul>
11. Information bias—differential effort by interviewer to identify exposures for cases and controls and their mothers (interviewers will not be blinded to case/control status).	<ul style="list-style-type: none"> <li>▶ Ensure that interviewers stick to the carefully constructed questionnaire (including preambles before questions) when conducting face-to-face and telephone interviews.</li> <li>▶ Have an online option for completing the study questionnaire, which cannot be influenced by this potential bias.</li> </ul>
12. Risks associated with retaining genome-wide genotyping data.	<ul style="list-style-type: none"> <li>▶ No biospecimens will be banked for the purposes of unspecified future research.</li> <li>▶ Genomic data for the current study will only ever be used to investigate the aetiology of TC, with the express consent of participants for using the data in this manner.</li> <li>▶ Genomic data will never leave New Zealand, but will only be analysed by the study team.</li> <li>▶ Control of the use of the genomic data collected for this study will rest with the Kaitiakitanga (Guardianship) Group, comprised senior Māori members of the study team, a cancer consumer representative and a representative from our Mother's Advisory Group.</li> </ul>

Continued

Table 1 Continued

Potential risk	Approach to risk management
13. Risks of stigmatising participants due to sensitive nature of some questions, for example, mother's use of cannabis use during pregnancy.	<ul style="list-style-type: none"> <li>▶ Data collection to be completed by CBG Health Research, a highly professional interviewing service who perform the NZ Health Survey for the Ministry of Health.</li> <li>▶ Online option for participants, to ensure privacy when answering questions and remove the need for a participant to say their answer aloud.</li> <li>▶ During face-to-face interviews, participants will be given a tablet to answer sensitive questions, to remove the need to say their answer aloud</li> <li>▶ All data analysis to be completed by a small, culturally competent research group (JKG, DS, JS).</li> <li>▶ Data interpretation to be assisted by wider study investigator group, who have substantial experience in competent data interpretation.</li> <li>▶ Finally, data interpretation will be completed by three separate groups: (1) senior Māori members of the study team; (2) a small cancer consumer group and (3) a Mother's Advisory Group.</li> </ul>
14. Risks of psychological distress due to sensitive nature of some questions, for example, mother's sense of self-blame when answering questions about exposures during pregnancy.	<ul style="list-style-type: none"> <li>▶ Data collection to be completed by CBG Health Research, a highly professional interviewing service who perform the NZ Health Survey for the Ministry of Health.</li> <li>▶ Participants will be reminded throughout the process (preinterview, during interview, after interview) that just because we are asking questions about an exposure does not mean that the exposure causes TC.</li> <li>▶ In the lay summary of study results, participants (specifically cases and their mothers) will be reassured that just because we may have found an association between a given exposure and the development of TC at a study level, that does not mean that their cancer was caused by that exposure.</li> <li>▶ Interview to be immediately ceased by the interviewer as soon as a participant becomes psychologically distressed. If interviewing in person or over the telephone, the interviewer will remain talking to the participant and reassuring them until they are calm and happy, at which point the interview will end.</li> <li>▶ The interviewer will also follow up the next day to ensure that the participant is feeling okay.</li> <li>▶ The participant would not be asked to take part in the study again (unless they expressed a strong personal desire to do so).</li> </ul>

### Genetic data

Logistic regression methods will be used to determine the odds of TC among those exposed to a given risk allele compared with non-exposed—both for the total cohort (ie, all cases/controls) and also stratified by ethnic group (Māori and non-Māori). From these analyses, ORs and their confidence limits will be calculated. In addition, we will describe the prevalence of given genetic risk alleles stratified by ethnic group. Both the individual associations between exposure to risk alleles and TC, as well as a quantification of the overall risk allele 'burden' (ie, the number of risk alleles presents in a given group), will also be determined.<sup>27 28</sup> For future analysis, we will use the same statistical approaches (with relevant adjustment for multiple comparisons) to test for the presence of additional susceptibility variants among Māori across the whole genome, with particular focus on previously identified loci (ie, genome-wide association study).

Gene/environment interaction will be assessed by combining interview data with genetic data. Given our sample size, this work will be exploratory in nature. First, we will stratify on the genetic exposure and assess the independent effect of the environmental exposure among those with and without the genetic exposure using logistic regression models. In addition, logistic regression models with an interaction term between relevant genetic and environmental exposures added to the baseline model will be fitted (ie, a model which includes the relevant

genetic and environmental exposure independently). We will use decision rules to identify the gene–environment interactions that will be explored, whereby we will only include genetic and environmental factors that are shown to be associated with TC development in the adjusted models. The presence of a gene/environment interaction will be assessed using likelihood ratio tests between the baseline model(s) and the model(s) which include the interaction term.<sup>29</sup> The detection of significant gene–environment interactions will only be possible where strong interactions are present, since very large sample sizes are required to detect weaker interactions with reasonable statistical precision.<sup>29</sup> As such, the basic stratified analysis—showing the distribution of the given environmental and given genetic exposure across cases and controls, with associated ORs—will form the primary results for this component of the study.

In summary, we will complete data analysis for the total combined sample of cases and controls, and also complete the same analysis stratified by ethnicity. For the combined analysis, we will (1) describe the prevalence of environmental and genetic exposures among TC cases and controls and (2) determine which exposures are important drivers of TC development in general. For the ethnicity-stratified analysis, we will (1) describe the prevalence of exposures among TC cases and controls, and whether this differs for Māori and non-Māori and (2) determine which exposures are important drivers of TC

development for Māori, and whether these differ to those that are important for non-Māori.

### Sample size and statistical power

#### Total analysis

In order for the case-control study to have sufficient statistical power to detect an OR of 1.6 (80% power with 95% CI) for reasonably common exposures (20%–70% underlying prevalence in control group), and greater ORs (95% CI 2.0 to 3.0) for rare (<20%) or very common (>70%) exposures, we require a sample size of 410 cases and 410 controls (Open Epi, V.3.0). As a realistic example of the likely statistical power of the study in terms of the genetic analysis, the risk allele frequency of the KITLG variant rs1508595 has been shown in TC genetics literature to be 83% among controls and 92% among TC cases, with an associated per-allele OR of 2.55 (95% CI 2.05 to 3.18).<sup>27</sup> The current study would be sufficiently powered to detect a difference in a prevalence of this magnitude (97% power, Open Epi, V.3.0). Adjustment for multiple comparisons (particularly when looking for evidence of relevant genetic variation across the whole genome) will reduce the power of the study from this nominal level.

#### Ethnicity-stratified analysis

Considering the focus of the study is to explain a disparity between Māori and non-Māori men, we aim to oversample the Māori population and recruit at least 150 Māori TC cases and 150 controls. In addition to the wider case-control study, oversampling Māori will allow us to stratify by ethnicity, and conduct separate nested case-control studies within ethnic groups (Māori stratified analysis: 150 cases vs 150 controls; non-Māori: 260 cases vs 260 controls). For Māori, a sample of 150 cases and 150 controls has 80% power to detect an OR of 2.0 for relatively common exposures (25%–60% prevalence in control group) and 2.0–3.0 for less (or more) common exposures. For non-Māori, a sample of 260 cases and 260 controls has 80% power to detect an OR of 1.7–1.8 for relatively common exposures (25%–70% prevalence in control group) and 2.0–3.0 for less (or more) common exposures. As an example of study power following stratification by ethnicity, we would have sufficient power (80%) to detect a difference if 40% of the mothers of Māori cases smoked during pregnancy compared with 25% of the mothers of Māori controls (OR 2.0).

#### Data management

A Kaitiakitanga (Guardianship) Group—comprised the senior Māori members of the study team, as well as a cancer consumer representative and a representative from our Mother's Advisory Group—will be created to provide guardianship over the data collected for this study. All interview and genomic data will be stored on password-protected computers. Once the final dataset is received by the study investigators, all identifiable information (including NHI number) will be deleted from the final master dataset. All data will be stored on a

password-protected computer, and only ever accessed by a small group of the study investigators. Anytime that data are transferred electronically, this data will be encrypted with a password. The data will be stored for a minimum of 10 years.

#### Future use of genomic data

Regarding the potential future use of genome-wide genotyping data for the purposes of conducting a genome-wide association study (to identify additional as-yet undiscovered predisposing genetic variants in Māori men, that is, research question 5 of the wider study), this would be tightly controlled by the study's Kaitiakitanga (Guardianship) Group. Our Kaitiakitanga (Guardianship) Group will need to unanimously approve any future use of genomic data prior to the study investigators seeking new ethical approval (or an amendment) to conduct further analysis. This will also be consented by the participants, and would be the subject of a new ethics application (or amendment to the existing study). This data would only ever be used to investigate the aetiology of TC.

#### Study timelines

The study protocol and questionnaire have been developed over the course of the past several years. Data collection for the pilot phase of the study began in May 2018, with associated (minor) revisions to the study protocol made following this pilot study phase. The purpose of this pilot phase is to answer questions regarding logistical operation of the study, including the following: (1) What are the difficulties associated with recruiting men into a TC case-control study? (2) How acceptable are the study questions to the cases/controls? (3) What are the difficulties associated with finding and interviewing their mothers, and what is the likely response rate for maternal questionnaires? and (4) What difficulties may be associated with the collection of oral swab or saliva samples, and what is the likely response rate for this component? Answering these questions will enable us to make minor adjustments to our study protocol (largely to maximise response rate).

We are currently seeking funding for the full study, and if successful within the 2018–2019 funding rounds we aim to have data collection, analysis and dissemination completed by the end of 2021.

#### Patient and public involvement

We are pleased to have the support of a small group of cancer consumers, who will consider the results from a cancer consumer's perspective prior to dissemination of study findings. Participants will be invited to receive a lay summary of the study results, along with an explanation of what these results might mean in the wider context of TC. This lay summary will begin by explaining in lay terms (aimed specifically at cases and their mothers) that the study-level findings do not necessarily mean that an individual's cancer was caused by the exposures that were found to be most strongly associated with TC. These



summaries will be prepared by the wider study investigator team, and screened by the senior Māori members of the team, the Mother's Advisory Group and our group of cancer consumers to ensure that they minimise the risk of participant stigmatisation.

## ETHICS AND DISSEMINATION

The study also underwent consultation with the Ngāi Tahu Research Consultation Committee (University of Otago), which is a committee that provides feedback on the study from an independent Indigenous Māori perspective.

### Data interpretation

All data analysis will be conducted by a small research team (JKG, DS, JS) comprised Māori and non-Māori researchers with a substantial history of conducting high-quality, culturally competent research (particularly in the area of health inequalities for Māori). Following data analysis, interpretation of the results will pass through multiple filters to ensure that the messages that are presented from the study are culturally competent and minimise the risk of stigmatising certain population groups (eg, Māori). First, the initial interpretation of the study results will be conducted by the wider study investigator team. Second, smaller groups will consider then consider the interpretation made by the wider study group from unique perspectives: (1) the senior Māori members of the study team will consider the impact of the findings on Māori, and how best to disseminate the study findings; a small group of cancer consumers will consider the results from a cancer consumer's perspective; and members of a Mother's Advisory Group (created specifically for this study) will be asked to assess the study investigator's interpretation of the study results, to ensure that these are presented in a manner that minimises the risk of stigmatisation of mothers (in general and specifically those involved in the study).

### Dissemination

The research team conducting this study is well placed to disseminate research findings to a wide and varied audience, ranging from general academia, community groups and clinical settings. Following the data interpretation process outlined above, the principal investigator will ensure that research findings are summarised in a clear, concise and culturally appropriate manner to all participants and their whānau, and that these summaries will also be more widely disseminated to the community. We will also develop press releases and make these available to media. The principal investigators and wider research team will participate in meetings and workshops which are relevant to the current study, and disseminate our findings where appropriate. At the conclusion of the study, results will be made available as both printed and online resources, as well as in academic and clinical peer-reviewed journals and at conferences.

## Study risks and sources of bias

We have identified potential risks associated with the execution of the study (including potential sources of bias and issues of participant safety), and also considered and documented our strategy for managing these risks. These risks and management strategies are shown in [table 1](#).

## CONCLUSIONS

The trends observed in New Zealand with respect to TC incidence may provide us with a unique opportunity to learn more about the currently obscure exposures that cause TC, by explaining why Māori have the highest rates of this disease in New Zealand. Using a case-control study design and techniques from epidemiology and genetics, our experienced research team aims to identify the key exposures in the development of TC in New Zealand, with a focus on those factors which might explain the incidence disparity between Māori and non-Māori men. By identifying these points of difference, we will increase international understanding of the key exposures which lead to the development of TC.<sup>3 14</sup>

Given the many exposures that have been studied, it is possible that the observed disparity is not likely to be the result of a single exposure, but rather the result of two or more risk factors that interact to confer increased TC risk to the Māori population. However, this cannot be answered until the completion of our study.

**Contributors** JKG is the principal investigator of the study. He is leading study concept and design, drafted the study protocol, and revised the protocol based on feedback. JS, KM, LR, CS, RE, TRM, BR, JK, MM and MAK assisted with study concept and design and provided critical revision of protocol drafts. DS provided mentoring and supervision to JKG, assisted with study concept and design and provided critical revision of protocol drafts.

**Funding** This work was supported by the Health Research Council of New Zealand (HRC), grant number 14/052.

**Competing interests** None declared.

**Patient consent** Not required.

**Ethics approval** Ethical approval for this study was sought and received from the New Zealand Ministry of Health's Health and Disability Ethics Committee (reference # 17/NTA/248). The current protocol manuscript is an abridged version of that which was approved by this committee.

**Provenance and peer review** Not commissioned; peer reviewed for ethical and funding approval prior to submission.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

1. Winter C, Albers P. Testicular germ cell tumors: pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol* 2011;7:43–53.
2. Huyghe E, Matsuda T, Thonneau P. Increasing incidence of testicular cancer worldwide: a review. *J Urol* 2003;170:5–11.
3. Sarfati D, Shaw C, Blakely T, et al. Ethnic and socioeconomic trends in testicular cancer. *International Journal of Cancer* 2010;128:1683–91.



4. Chia VM, Quraishi SM, Devesa SS, *et al.* International trends in the incidence of testicular cancer, 1973-2002. *Cancer Epidemiol Biomarkers Prev* 2010;19:1151-9.
5. Gurney JK, Sarfati D, Stanley J. Obscure etiology, unusual disparity: the epidemiology of testicular cancer in New Zealand. *Cancer Causes Control* 2015;26:561-9.
6. McGlynn KA, Cook MB. *Male Reproductive Cancers: Epidemiology, Pathology and Genetics*: Springer-Verlag, 2010.
7. Shah MN, Devesa SS, Zhu K, *et al.* Trends in testicular germ cell tumours by ethnic group in the United States. *Int J Androl* 2007;30:206-13.
8. McGlynn KA, Devesa SS, Sigurdson AJ, *et al.* Trends in the incidence of testicular germ cell tumors in the United States. *Cancer* 2003;97:63-70.
9. Garner MJ, Turner MC, Ghadirian P, *et al.* Epidemiology of testicular cancer: an overview. *Int J Cancer* 2005;116:331-9.
10. Maule M, Malavassi JL, Richiardi L. Age at puberty and risk of testicular cancer: a meta-analysis. *Int J Androl* 2012;35:828-34.
11. Brown LM, Potters LM, Hoover RN. Testicular cancer in young men: the search for causes of the epidemic increase in the United States. *J Epidemiol Community Health* 1987;41:349-54.
12. Akre O, Pettersson A, Richiardi L. Risk of contralateral testicular cancer among men with unilaterally undescended testis: a meta analysis. *Int J Cancer* 2009;124:687-9.
13. Lip SZ, Murchison LE, Cullis PS, *et al.* A meta-analysis of the risk of boys with isolated cryptorchidism developing testicular cancer in later life. *Arch Dis Child* 2013;98:20-6.
14. Gurney J, Sarfati D, Stanley J, *et al.* Do ethnic patterns in cryptorchidism reflect those found in testicular cancer? *J Urol* 2013;190:1852-7.
15. Gurney JK, McGlynn KA, Stanley J, *et al.* Risk factors for cryptorchidism. *Nat Rev Urol* 2017;14:534-48. Advance online publication.
16. Richiardi L, Vizzini L, Pastore G, *et al.* Lifetime growth and risk of testicular cancer. *Int J Cancer* 2014;135:695-701.
17. McGlynn KA, Sakoda LC, Rubertone MV, *et al.* Body size, dairy consumption, puberty, and risk of testicular germ cell tumors. *Am J Epidemiol* 2007;165:355-63.
18. Béranger R, Pérol O, Bujan L, *et al.* Studying the impact of early life exposures to pesticides on the risk of testicular germ cell tumors during adulthood (TESTIS project): study protocol. *BMC Cancer* 2014;14:563.
19. Ministry of Health. *New Zealand Health Survey Methodology Report 2015/16*. Wellington, New Zealand: Ministry of Health, 2016.
20. Ministry of Health. *New Zealand Health Information Service*. Wellington, New Zealand, 2004.
21. *Ministry of Health*. Wellington: Ministry of Health, 2009:5.3.
22. Ministry of Health. *New cancer registrations 2014, 2016*.
23. Chung CC, Kanetsky PA, Wang Z, *et al.* Meta-analysis identifies four new loci associated with testicular germ cell tumor. *Nat Genet* 2013;45.
24. Ruark E, Seal S, McDonald H, *et al.* Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. *Nat Genet* 2013;45:686-9.
25. Litchfield K, Levy M, Orlando G, *et al.* Identification of 19 new risk loci and potential regulatory mechanisms influencing susceptibility to testicular germ cell tumor. *Nat Genet* 2017;49:1133-40.
26. Wang Z, McGlynn KA, Rajpert-De Meyts E, *et al.* Meta-analysis of five genome-wide association studies identifies multiple new loci associated with testicular germ cell tumor. *Nat Genet* 2017;49:1141-6.
27. Rapley EA, Turnbull C, Al Olama AA, *et al.* A genome-wide association study of testicular germ cell tumor. *Nat Genet* 2009;41:807-10.
28. Lewis CM, Vassos E. Prospects for using risk scores in polygenic medicine. *Genome Med* 2017;9:96.
29. Nickels S, Truong T, Hein R, *et al.* Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. *PLoS Genet* 2013;9.