Assessing generalisability through the use of disease registers: findings from a diabetes cohort study

Michael David, Robert Ware, Maria Donald, Rosa Alati

ABSTRACT

Objectives: Knowledge of a study population’s similarity to the target population allows researchers to assess the generalisability of their results. Often generalisability is assessed through a comparison of baseline characteristics between individuals who did and did not respond to an invitation to participate in a study. In this prospective population-based cohort, we broadened this assessment by comparing participants with all individuals from a chronic disease register who satisfied the study eligibility criteria but for a number of reasons, such as the absence of consent to be approached for research purposes, did not participate.

Methods: Data are from the Living with Diabetes Study, a population-based cohort of individuals diagnosed with diabetes mellitus, which commenced in Queensland, Australia in 2008. Individuals were sampled from a federally-funded diabetes register. We compared the characteristics of 3951 study participants with 10 488 non-participants (individuals who were invited to participate but declined) and with 129 900 non-study individuals on the register who did not participate in the study.

Results: Study participants were more likely than non-study registrants to be male, aged 50–69, have type 2 diabetes non-insulin requiring, be recently registered and be non-Indigenous Australians. Study participants were more likely than non-participants to be aged 50–69, have type 1 diabetes and be non-Indigenous Australians.

Conclusions: The interpretation of a study’s generalisability can alter depending on which non-participating group is compared with participants. When assessing generalisability, participants should be compared with the largest possible group of non-participating individuals. When sampling from a disease register, researchers should be wary of the influence of research consent procedures on the register’s coverage.

INTRODUCTION

Population-based cohort studies are essential when studying chronic diseases such as diabetes mellitus, as they can offer a comprehensive understanding of disease trajectory over time and allow for multiple subgroup analyses. However, the utility of each study’s findings depends on whether the results are sufficiently generalisable to the population under investigation. The extent of a study’s generalisability, or external validity, depends on how representative of the target population the study’s participants are. Since information on the target population is often unavailable, investigations concerning the generalisability of population-based cohorts, including those concerning diabetes, have focused on the comparison of baseline characteristics between study participants and non-participants to assess how similar or different they
are.

However, a high degree of similarity between participants and non-participants does not necessarily mean results arising from the study will have good generalisability, as these two groups as a whole might not fully represent the target population due to coverage error.


The extent to which findings are generalisable can be assessed by comparing the study participants with the largest possible subset of all diseased individuals in the population being studied, other than participants. The characteristics of this larger group can be accessed through databases such as national chronic disease registers. Chronic disease registries are increasingly used to recruit participants to cohort studies. One purported advantage of this is to ensure generalisability to the target population. In recent years, however, legislative reform concerning privacy issues has been introduced in many countries, including Australia, which has restricted research related access to these databases without an individual’s consent. It is possible that this may seriously limit the usefulness of chronic disease registers for epidemiological research.

This study investigates the generalisability of one Australian register, the National Diabetes Services Scheme (NDSS) and explores whether chronic disease registrants who agreed to participate in a research study have similar characteristics to registrants who satisfied the inclusion criteria but did not participate. We also compare the characteristics of participants with the characteristics of individuals who were invited to participate but declined.

METHODS

The Living with Diabetes Study (LWDS) is a population-based cohort study that began in Queensland, Australia in 2008. An individual was eligible to participate in the study if they had doctor diagnosed type 1 or 2 diabetes, were aged at least 18 years and had a valid Queensland postal address. Individuals were randomly sampled from a federally-funded register of Australians with diabetes, the NDSS, managed by a non-governmental organisation called Diabetes Australia. The NDSS’s coverage of Queenslanders with diabetes is estimated to be between 80% and 90%. Since 2001, individuals joining the NDSS have been asked whether they would like to be informed about opportunities to participate in research. Only those who consented to be contacted for research purposes were eligible to be invited to participate in the LWDS. The LWDS sampling design specified three target locations of policy interest to be oversampled: an outer metropolitan area, a new suburban development and a coastal agricultural community. All eligible individuals from the three locations were invited to participate; approximately one in six eligible individuals from the rest of Queensland were invited to participate.

Selected individuals were invited to participate in the LWDS via a mailed questionnaire. Information was collected on demographic and socio-economic characteristics, health behaviour, and health and psychological status. Strategies to maximise participation included reminder cards, telephone calls and replacement surveys. We categorised registrants into four mutually exclusive groups: participants, non-participants, non-sampled consenting registrants and non-consenting registrants. A participant was defined as an individual who agreed to participate in the LWDS. A non-participant was defined as an individual who was invited to participate in the LWDS but declined. A non-sampled consenting registrant was defined as an individual who agreed to participate in the LWDS but was not selected during the sampling process. A non-consenting registrant was defined as an individual who had not agreed to be contacted for research purposes. Initially, participants were compared to non-participants (the reference group). For a secondary comparative analysis, the reference group was expanded to comprise all registrants who were not study participants. This expanded reference group was defined as non-study registrants (figure 1).

Data analysis

For participants, non-participants and non-study registrants, we calculated the frequency (percentage) of individuals in each category for sex, age (18–49, 50–69, 70+ years), diabetes status (type 2 non-insulin requiring; 16 In recent years, however, legislative reform concerning privacy issues has been introduced in many countries, including Australia, which has restricted research related access to these databases without an individual’s consent. It is possible that this may seriously limit the usefulness of chronic disease registers for epidemiological research.

This study investigates the generalisability of one Australian register, the National Diabetes Services Scheme (NDSS) and explores whether chronic disease registrants who agreed to participate in a research study have similar characteristics to registrants who satisfied the inclusion criteria but did not participate. We also compare the characteristics of participants with the characteristics of individuals who were invited to participate but declined.

METHODS

The Living with Diabetes Study (LWDS) is a population-based cohort study that began in Queensland, Australia in 2008. An individual was eligible to participate in the study if they had doctor diagnosed type 1 or 2 diabetes, were aged at least 18 years and had a valid Queensland postal address. Individuals were randomly sampled from a federally-funded register of Australians with diabetes, the NDSS, managed by a non-governmental organisation called Diabetes Australia. The NDSS’s coverage of Queenslanders with diabetes is estimated to be between 80% and 90%. Since 2001, individuals joining the NDSS have been asked whether they would like to be informed about opportunities to participate in research. Only those who consented to be contacted for research purposes were eligible to be invited to participate in the LWDS. The LWDS sampling design specified three target locations of policy interest to be oversampled: an outer metropolitan area, a new suburban development and a coastal agricultural community. All eligible individuals from the three locations were invited to participate; approximately one in six eligible individuals from the rest of Queensland were invited to participate.

Selected individuals were invited to participate in the LWDS via a mailed questionnaire. Information was collected on demographic and socio-economic characteristics, health behaviour, and health and psychological status. Strategies to maximise participation included reminder cards, telephone calls and replacement surveys. We categorised registrants into four mutually exclusive groups: participants, non-participants, non-sampled consenting registrants and non-consenting registrants. A participant was defined as an individual who agreed to participate in the LWDS. A non-participant was defined as an individual who was invited to participate in the LWDS but declined. A non-sampled consenting registrant was defined as an individual who agreed to participate in the LWDS but was not selected during the sampling process. A non-consenting registrant was defined as an individual who had not agreed to be contacted for research purposes. Initially, participants were compared to non-participants (the reference group). For a secondary comparative analysis, the reference group was expanded to comprise all registrants who were not study participants. This expanded reference group was defined as non-study registrants (figure 1).

Data analysis

For participants, non-participants and non-study registrants, we calculated the frequency (percentage) of individuals in each category for sex, age (18–49, 50–69, 70+ years), diabetes status (type 2 non-insulin requiring;
type 2 insulin requiring, type 1), registration year (2001–2003, 2004–2005, 2006–2008), SEIFA tertile and indigenous status. Initially, we used logistic regression analyses to compare participants with non-participants on a univariable basis. We then fitted a series of multivariable logistic regression models in order to investigate the impact of potential confounders and obtain fully adjusted associations. Analyses were weighted according to the sampling scheme. As individual-level data were not available for all individuals in the reference group of non-study registrants, we used univariable logistic regression with aggregate data to compare participants with non-study registrants. Results for each of the analyses are presented in table 1 as ORs and 95% CIs.

RESULTS
On 30 June 2008 there were 133 851 registrants in the NDSS who satisfied the LWDS entry criteria, of whom 75 347 (56.3%) did not consent to participate in any registration year and SEIFA. On 30 June 2008 there were 133 851 registrants in the LWDS, 3951 of whom agreed. Complete aggregate information was available for all variables except for registration year and SEIFA. Due to NDSS procedural changes and invalid postcodes, data for 56 264 registration years and 1711 postcodes were not available for the analyses.

Table 1 shows the results of comparisons of 3951 participants with 10 488 non-participants, and of participants with 129 900 non-study registrants. After adjusting for all covariates, individuals were less likely to participate in the LWDS if they were younger (OR 0.63; 95% CI 0.55 to 0.71) or older (0.89; 0.81 to 0.99) than those aged 50–69 years and had identified themselves as being indigenous Australians (0.61; 0.48 to 0.77). Those who had type 1 diabetes (1.50; 1.19 to 1.90) were more likely to participate in the study. A sensitivity analysis was conducted to specifically investigate the effect of potential confounders. The analyses were re-run six times with one covariate excluded on each occasion. The only effect estimate seen to vary substantially was diabetes status. In the model adjusted across all covariates, individuals were less likely to participate in the LWDS if they were insulin requiring, and 72% less likely for those with type 2 diabetes who were insulin requiring, with the odds of participation being 29% less likely for those with type 2 diabetes who were insulin requiring, and 72% less likely for those with type 1 diabetes. This parallels the research literature, which suggests that those less healthy are more likely to be non-responders than those in better health. However, this was not the case when individuals were compared to non-participants, which showed a strong association also, but was directionally opposite to the previous result; the adjusted odds of those with type 1 diabetes participating were 50% greater than those who had type 2 diabetes but were not insulin requiring. Such a result indicates that those with type 1 diabetes, although less likely to be invited due to consent issues relating to age at diagnosis, were more likely to participate once invited.

Age and Australian indigenous status were also significantly associated with study participation, with age also having a negative confounding effect on the LWDS participation—diabetes status relationship. Unlike the influence of diabetes status, these associations were similar in direction and strength for both comparative analyses.

Although these results are consistent with the literature, they raise the issue of representativeness. Disparities in sample balance have the potential to impact adversely on the estimation of population parameters such as prevalence and incidence metrics. Our initial comparative analysis was between participants and non-participants, and relied solely on information from those invited to participate in the study.
<table>
<thead>
<tr>
<th>Sex</th>
<th>Participants N = 3951</th>
<th>Non-participants N = 10,488</th>
<th>Non-study registrants N = 129,900</th>
<th>Participants versus non-participants</th>
<th>Participants versus non-study registrants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Crude OR (95% CI)</td>
<td>Adjusted OR (95% CI)*</td>
<td>Crude OR (95% CI)</td>
<td></td>
<td>Crude OR (95% CI)</td>
</tr>
<tr>
<td>Male</td>
<td>2176 (55.1%)</td>
<td>5885 (56.1%)</td>
<td>68,618 (52.8%)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Female</td>
<td>1775 (44.9%)</td>
<td>4603 (43.9%)</td>
<td>61,282 (47.2%)</td>
<td>1.01 (0.93 to 1.10)</td>
<td>1.07 (0.98 to 1.17)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–49</td>
<td>618 (15.6%)</td>
<td>246 (21.4%)</td>
<td>21,387 (16.5%)</td>
<td>0.67 (0.60 to 0.75)</td>
<td>0.63 (0.55 to 0.71)</td>
</tr>
<tr>
<td>50–69</td>
<td>2375 (60.1%)</td>
<td>5649 (53.9%)</td>
<td>58,988 (45.4%)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>70+</td>
<td>958 (24.3%)</td>
<td>2593 (24.7%)</td>
<td>49,525 (38.1%)</td>
<td>0.88 (0.80 to 0.97)</td>
<td>0.89 (0.80 to 0.99)</td>
</tr>
<tr>
<td>Diabetes status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 2, no insulin</td>
<td>3023 (76.5%)</td>
<td>8024 (76.5%)</td>
<td>82,717 (63.7%)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Type 2, insulin</td>
<td>738 (18.7%)</td>
<td>1986 (18.9%)</td>
<td>28,336 (21.8%)</td>
<td>0.97 (0.87 to 1.08)</td>
<td>0.97 (0.86 to 1.11)</td>
</tr>
<tr>
<td>Type 1, insulin</td>
<td>190 (4.8%)</td>
<td>478 (4.6%)</td>
<td>18,847 (14.5%)</td>
<td>1.11 (0.91 to 1.34)</td>
<td>1.50 (1.19 to 1.90)</td>
</tr>
<tr>
<td>Registration year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001–2003</td>
<td>1303 (38.0%)</td>
<td>3422 (37.0%)</td>
<td>28,714 (38.8%)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2004–2005</td>
<td>805 (23.4%)</td>
<td>2239 (24.2%)</td>
<td>20,024 (27.0%)</td>
<td>0.96 (0.85 to 1.07)</td>
<td>0.97 (0.87 to 1.09)</td>
</tr>
<tr>
<td>2006–2008</td>
<td>1325 (38.6%)</td>
<td>3580 (38.8%)</td>
<td>25,389 (34.2%)</td>
<td>1.01 (0.92 to 1.12)</td>
<td>1.07 (0.97 to 1.19)</td>
</tr>
<tr>
<td>SEIFA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>830 (21.0%)</td>
<td>2491 (23.8%)</td>
<td>27,049 (21.1%)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Middle</td>
<td>1543 (39.1%)</td>
<td>3883 (37.1%)</td>
<td>51,932 (40.5%)</td>
<td>1.13 (1.01 to 1.27)</td>
<td>1.11 (0.98 to 1.25)</td>
</tr>
<tr>
<td>High</td>
<td>1572 (39.9%)</td>
<td>4100 (39.1%)</td>
<td>49,214 (38.4%)</td>
<td>1.11 (0.99 to 1.24)</td>
<td>1.07 (0.95 to 1.21)</td>
</tr>
<tr>
<td>Indigenous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>3838 (97.2%)</td>
<td>9969 (95.1%)</td>
<td>124,033 (95.5%)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Yes</td>
<td>113 (2.8%)</td>
<td>519 (4.9%)</td>
<td>5867 (4.5%)</td>
<td>0.57 (0.45 to 0.71)</td>
<td>0.61 (0.48 to 0.77)</td>
</tr>
</tbody>
</table>

*Adjusted for: sex, age category, diabetes status, registration year, SEIFA status and indigenous status.

SEIFA, Socio-Economic Indexes for Areas (the Australian Bureau of Statistics’ index of relative socio-economic disadvantage).
This analysis failed to identify an important association between diabetes status and participation. This was due to the underrepresentation of individuals with type 1 diabetes by a factor of more than 3 in the group of non-participants (4.6%) when compared to registrants not invited to participate in the study (15.4%). Such underrepresentation is the consequence of type 1 diabetes being predominately diagnosed during childhood and the NDSS consent protocol, which does not include a systematic updating of consent status at the age of 18 among those registered as a child.

Mandatory informed consent, including parental, not only has a negative effect on participation rates overall, but also weakens the representativeness of the study sample by producing unbalanced subgroups among the study participants. This was the case because research consent was not a necessary criterion for an individual to be considered a registrant.

The results of our study should be interpreted within the context of some limitations. First, the generalisability of any study’s findings to the target population is very much dependent on register coverage and the quality of its database. Increased levels of coverage and data quality lessen the likelihood of biased sample estimates. The coverage of the NDSS is estimated to be between 80% and 90%, which is higher than most diabetes registers, thus giving it the potential to produce sampling frames of a higher data quality than

---

**Figure 2** Participation flowchart for the Living with Diabetes Study.
most. Second, in analyses such as these which only utilise one time-point, there is an inability to maximise the information provided by time varying determinants of non-response such as age. Third, due to unavailability of individual-level data for registrants not invited to participate in the study, it was not possible to complete a comparative analysis between participants and non-study registrants that isolated the independent covariate effects after adjustment. It is possible that aggregate data would have resulted in the associations between participation and a number of covariates being more similar to those found when non-participants were used as the reference group.

Our findings illustrate that the standard procedure of comparing study participants and non-participants in assessing a study’s generalisability can be compromised by the issue of research consent when disease registers are used as a source of recruitment. Whenever possible, a clearer assessment should be sought by extending this standard practice to a secondary analysis by sourcing the largest possible reference group that is inclusive of non-participants. For prospective population-based cohort studies, researchers should endeavour to source a group that contains all potential participants who satisfied inclusion criteria but have not been able to participate. As findings can be influenced by the issue of research consent, where available, chronic disease registers should be utilised fully in any assessment of generalisability.

Acknowledgements We would like to especially thank the participants of the Living with Diabetes Study; without their participation this research would not have been possible. Also, our sincere thanks go to Diabetes Australia and the National Diabetes Services Scheme for working with us to make it possible to recruit participants to the Living with Diabetes Study. In addition, we would also like to thank all members of the Living with Diabetes Study team for their ongoing support and input.

Funding This research was funded by the Australian Research Council (DP0988805) and Queensland Health through the Queensland Strategy for Chronic Disease 2005–2015.

Competing interests None.

Ethics approval Ethics approval was provided by The University of Queensland’s Behavioural and Social Sciences Ethical Review Committee (BSSERC).

Contributors All authors designed the study. M Donald was responsible for data acquisition. M David and R Ware analysed the data. M David drafted the initial manuscript. R Ware, M Donald and R Alati critically reviewed the manuscript. All authors read and approved the final manuscript.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data used in this study were de-identified (personal identifiers were removed or disguised) before usage. These data are organised and held at the School of Population Health at The University of Queensland. The Living with Diabetes Study’s homepage includes information on methodology and study updates including preliminary descriptive findings (www.lwds.org.au). Researchers are invited to contact the research team to explore options for collaborative work.

REFERENCES