BMJ open

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

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2011-000078
Article Type:	Research
Date Submitted by the Author:	18-Feb-2011
Complete List of Authors:	David, Michael; The University of Queensland, School of Population Health Ware, Robert; The University of Queensland, School of Population Health Donald, Maria; The University of Queensland, School of Population Health Alati, Rosa; The University of Queensland, School of Population Health
Subject Heading:	Statistics & research methods
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS, General diabetes < DIABETES & ENDOCRINOLOGY

TITLE PAGE

TITLE:

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

CORRESPONDING AUTOR:

Name: Michael David
Postal Address: School of Population Health, The University of Queensland, Herston, Queensland 4000, Australia
E-mail: michael.david@uqconnect.edu.au
Tel: +61 733655298
Fax: +61 733655298

CO-AUTORS:

Robert Ware[1]; Maria Donald[1]; and Rosa Alati[1]
[1] School of Population Health, The University of Queensland, Herston, Queensland, Australia

KEY WORDS and PHRASES:

Generalizability; Cohort Study; Research Consent; Disease Register

WORD COUNT:

2,109

ABSTRACT

Objectives

The knowledge of a study population's similarity to the target population allows researchers to assess the generalizability of their results. Often generalizability 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 3,951 study participants with 10,488 non-participants (individuals who were invited to participate but declined), and with 129,900 non-study registrants (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 to 69, have Type 2 diabetes not requiring insulin, and be non-indigenous Australians. Study participants were more likely than non-participants to be female, aged 50 to 69 have Type 1 diabetes, have higher socio-economic status, and be non-indigenous Australians.

Conclusions

The interpretation of a study's generalizability can alter depending on which non-participating group is compared with participants. When assessing generalizability, 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.

Article Summary

Article Focus

* to assess the similarity between participants recruited into a study who were sampled from a chronic disease register, with registrants who did not participate in the study
* to assess whether the differences between participants and registrants who did not participate are similar to the differences between participants and individuals who were invited to participate but declined

Key Messages

* the generalizability of a study should be assessed by comparing participants 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

Strengths and Limitations

* information is available for all individuals registered with the chronic disease register
* the chronic disease register from which study participants were recruited from has high coverage of target population
* only aggregated data was available for registrants who were not invited to participate in the study

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.[1, 2] However the utility of each study's findings depends on whether the results are sufficiently generalizable to the population under investigation. The extent of a study's generalizability, or external validity, depends on how representative of the target population the study's participants are.[3-5] Since information on the target population is often unavailable, investigations concerning the generalizability 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.[6-9] However, a high degree of similarity between participants and non-participants does not necessarily mean results arising from the study will have good generalizability, as these two groups as a whole might not fully represent the target population due to coverage error.[10, 11] The extent to which findings are generalizable 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.[12, 13] The characteristics of this larger group can be accessed through databases such as national chronic disease registers.[6]

Chronic disease registries are increasingly used to recruit participants to cohort studies. One purported advantage of this is to ensure generalizability to the target population.[14-16] In recent years however, legislative reform concerning privacy issues has been introduced in many countries, including Australia,[17] 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 epidemiologic research.[18, 19] This study investigates the generalizability 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 2008. It is an annual study conducted in Queensland, Australia. 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 named Diabetes Australia.

The NDSS's coverage of Queenslanders with diabetes is estimated to be between 80% and 90%.[20] Since 2001, individuals joining the NDSS have been asked whether they would like to be informed about opportunities to participate in research. Those who consented to be contacted for research purposes and had a valid postal address were invited to participate in the LWDS.

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 maximize participation included reminder cards, telephone calls and replacement surveys. We categorized registrants into three mutually exclusive groups: participants, non-participants, and non-invitees. 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-invitee was defined as an individual who was not invited to enrol in the LWDS, either because they did not consent to being approached by research teams, or because they were not selected during the sampling process. For comparative purposes, these two latter groups were combined and defined as non-study registrants. Available individual-level information on participants and non-participants consisted of sex, age, diabetes status, year of NDSS registration, postcode and indigenous status. Postcodes were matched to the Australian Bureau of Statistics' Index of Relative Socio-economic Disadvantage (SEIFA) ranking, and categorised into tertiles.[21] Only aggregated data was available for non-invitees.

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 and 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 multivariate logistic regression models in order to investigate the impact of potential confounders and obtain fully-adjusted associations. Individuals were weighted according to the sampling scheme. Finally, we used logistic regression with aggregated data to compare participants with non-study registrants. Results are presented in Table 1 as odds ratios (ORs) and 95% confidence intervals (95% CIs).

RESULTS

At the 30th of 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 research and were excluded (Figure 1). Of the remaining 58,504 registrants, 14,439 were invited to participate in the LWDS, 3,951 of whom agreed. Complete aggregated information was available for all variables except for registration year and SEIFA. Due to NDSS procedural changes and invalid postcodes, 56,264 registrations and 1,711 postcodes were not available for the analyses.

[Figure 1 to be inserted here]

Table 1 displays a comparison of 3,951 participants and 10,488 non-participants, and a comparison between participants and 129,900 non-study registrants. After adjusting for all covariates, individuals were less likely to participate in the LWDS if they were younger (OR=0.62; 95% CI: 0.55-0.69) or older (0.89; 0.81-0.98) than those aged 50 to 69 years; and had identified themselves as being indigenous Australians (0.61; 0.49-0.76). Those who were female (1.10; 1.02-1.19); had Type 1 diabetes (1.43; 1.16-1.76) and resided in middle (1.16; 1.04-1.28) or high SEIFA areas (1.11; 1.01-1.24) 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 except age, the odds of being a participant if an individual had Type 1 diabetes was 1.06 (0.87 –1.29) greater than if an individual had Type 2 diabetes and not insulin requiring, while for the fully adjusted model it was 1.43 (1.16–1.76).

The comparative analyses between participants and non-study registrants (Table 1) shows a number of associational differences when compared to the previous multivariate analysis. The most noticeable is the relationship between participation and diabetes status, as it varies not only in strength, but direction. These analyses show that compared with Type 2 diabetes, no insulin, individuals with Type 2 diabetes, insulin dependent (0.71; 0.66-0.77), and Type 1 diabetes (0.28; 0.24-0.32) were less likely to participate. Similarly, the association between participation status

Table 1: Comparison between Living with Diabetes Participants (N=3,951), Non-participants (N=10,488) and Non-study registrants (N=129,900)

	Participants	Non-Participants	Non-study Registrants	Participants Vs Non-Participants		Participants vs Non-study Registrants
				Crude OR (95% CI)	Adjusted OR (95% CI)	Crude OR (95% CI)
	N (%)	N (%)	N (%)			
Sex						
Male	2,176 (55.1%)	5,885 (56.1%)	68,618 (52.8%)	1.00	1.00	1.00
Female	1,775 (44.9%)	4,603 (43.9%)	61,282 (47.2%)	1.04 (0.97 – 1.12)	1.10 (1.02 – 1.19)	0.91 (0.86 – 0.97)
Age						
18-49	618 (15.6%)	2,246 (21.4%)	21,387 (16.5%)	0.65 (0.59 - 0.72)	0.62 (0.55 – 0.69)	0.71 (0.66 – 0.79)
50 - 69	2,375 (60.1%)	5,649 (53.9%)	58,988 (45.4%)	1.00	1.00	1.00
70+	958 (24.3%)	2,593 (24.7%)	49,525 (38.1%)	0.88 (0.80 – 0.96)	0.89 (0.81 -0.98)	0.48 (0.45 – 0.52)
Diabetes Status						
Type 2 , No Insulin	3,023 (76.5%)	8,024 (76.5%)	82,717 (63.7%)	1.00	1.00	1.00
Type 2, Insulin	738 (18.7%)	1,986 (18.9%)	28,336 (21.8%)	0.99 (0.90 – 1.08)	1.00 (0.90 – 1.12)	0.71 (0.66 – 0.77)
Type 1, Insulin	190 (4.8%)	478 (4.6%)	18,847 (14.5%)	1.06 (0.89 – 1.25)	1.43 (1.16 – 1.76)	0.28 (0.24 – 0.32)
Registration Year						
2001 - 2003	1,303 (38.0%)	3,422 (37.0%)	28,741 (38.8%)	1.00	1.00	1.00
2004 - 2005	805 (23.4%)	2,239 (24.2%)	20,024 (27.0%)	0.94 (0.85 – 1.05)	0.96 (0.87 – 1.07)	0.89 (0.81 – 0.97)
2006 - 2008	1,325 (38.6%)	3,580 (38.8%)	25,389 (34.2%)	0.97 (0.89 – 1.06)	1.02 (0.94 – 1.13)	1.15 (1.06 – 1.24)
SEIFA						
Low	830 (21.0%)	**2**,491 (23.8%)	27,049 (21.1%)	1.00	1.00	1.00
Middle	1,543 (39.1%)	3,883 (37.1%)	51,932 (40.5%)	1.19 (1.08 – 1.32)	1.16 (1.04 – 1.28)	0.97 (0.89 – 1.05)
High	1,572 (39.9%)	4,100 (39.1%)	49,214 (38.4%)	1.15 (1.04 – 1.27)	1.11 (1.01 – 1.24)	1.04 (0.96 – 1.13)
Indigenous						
No	3,838 (97.2%)	9,969 (95.1%)	124,033 (95.5%)	1.00	1.00	1.00
Yes	113 (2.8%)	519 (4.9%)	5,867 (4.5%)	0.57 (0.46 – 0.70)	0.61 (0.49 – 0.76)	0.62 (0.52 – 0.75)

and sex was also reversed, with females less likely than males to be participants (0.91; 0.86-0.97). There was no evidence of an association between participation and SEIFA, but this was not the case with Year of NDSS registration, as registration between 2006 and 2008 was positively associated with participation (1.15; 1.06-1.24), whilst registration between 2004 and 2005 was inversely associated (0.89; 0.81-0.97). Associations between participation, and the covariates of age and indigenous status were similar in direction and magnitude with those found by the multivariate analysis, except for those aged at least 70 years, which strengthened inversely (0.48; 0.45-0.52).

DISCUSSION

The differences observed between the comparisons between participants and non-participants, and between participants and non-study registrants, confirm that the extent of a study's generalizability should be established by comparing study participants to a group of individuals which best represents the target population. In this study, those who agreed to participate in the LWDS were significantly different from the non-study registrants over a number of characteristics, with the most notable being diabetes status. Those with Type 2 diabetes who were insulin requiring, were less likely to participate in the LWDS. Individuals were less likely to be participants if they were insulin requiring, with the odds of participation being 29% less likely for those with Type 2 diabetes who were not 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 nonresponders than those in better health.[22-24] However, this was not the case when participants 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 43% greater than those who had Type 2 diabetes but were not insulin requiring. Such a result indicates that those with Type 1 diabetes, though less likely to be invited due to consent issues relating to age of diagnosis,[25] 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. Though these results are consistent with the literature,[5, 26, 27] 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.[9, 28-30]

Our initial comparative analysis was between participants and non-participants, and relied solely on information from those invited to participate in the study. 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 three in the group of invitees when compared to the non-study registrants. Such underrepresentation is the consequence of Type 1 diabetes being predominately diagnosed during childhood and the NDSS consent protocol,[20] which does not include a systematic updating of consent status at the age of 18 amongst 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 amongst the study participants.[25, 31, 32] 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. Firstly, the generalizability of any study's findings to the target population is very much dependent on register coverage and the quality of its database.[16, 33, 34] Increased levels of coverage and data quality lessen the likelihood of biased sample estimates. [35-37] The coverage of the NDSS is estimated to be between 80% and 90%, which is higher than most diabetes registers,[20, 33] thus giving it the potential to produce sampling frames of a higher data quality than most. Secondly, in analyses such as these

which only utilize one time-point, there is an inability to maximize the information provided by time varying determinants of nonresponse such as age.[23, 38, 39] Thirdly, due to unavailability of individual-level data for non-invitees from the NDSS, 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 individual 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 generalizability 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 utilized fully in any assessment of generalizability.

ACKNOWLEDGMENTS

We would like to especially thank the participants of the Living with Diabetes Study, without their participation this research would not be possible. Also, our sincere thanks go to Diabetes Australia and the National Diabetes 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.

COMPETING INTERESTS

There are no conflicts of interest with respect to this study as outlined in this paper.

CONTRIBUTORSHIP STATEMENT

I declare that I conceived the study and was the primary author responsible for this final version. In addition, Dr. Robert Ware assisted and advised on conceptualisation, statistical analysis and reviewing, while Dr. Rosa Alati and Dr. Maria Donald assisted in the review and editing of the final version of this paper.

FUNDING

We are grateful to both the Australian Research Council and Queensland Health for providing funding that enabled this study to be commenced at the data collection stage and finalized by the completion of this paper.

REFERENCES

1. Tabák AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimäki M, Witte DR. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. The Lancet. 2009;373(9682):2215-21.
2. Mrena S, Virtanen SM, Laippala P, Kulmala P, Hannila ML, Akerblom HK, et al. Models for predicting type 1 diabetes in siblings of affected children. Diabetes care. 2006;29(3):662-7.
3. Szklo M. Population-based cohort studies. Epidemiologic reviews. 1998;20(1):81.
4. Deeg D. Attrition in longitudinal population studies: Does it affect the generalizability of the findings? An introduction to the series. Journal of clinical epidemiology. 2002;55(3):213-5.
5. Drivsholm T, Eplov L, Davidsen M, Jørgensen T, Ibsen H, Hollnagel H, et al. Representativeness in population-based studies: a detailed description of non-response in a Danish cohort study. Scandinavian journal of public health. 2006;34(6):623-31.
6. Sim J. The external validity of group comparative and single system studies. Physiotherapy. 1995;81(5):263-70.
7. Gerrish K, Lacey A, editors. The Research Process in Nursing. 6th Edition ed. Chichester: John Wiley and Sons; 2010.
8. Barry A. How attrition impacts the internal and external validity of longitudinal research. Journal of School Health. 2005;75(7):4.

9. Livingston PM, Lee SE, McCarty CA, Taylor HR. A comparison of participants with non-participants in a population-based epidemiologic study: The Melbourne Visual Impairment Project. Ophthalmic epidemiology. 1997;42(2):73-81.

10. Groves R, Dillman D, Eltinge J, Little R, Biemer P, Lyberg L, et al. Survey methodology. Technometrics. 2005;47(2):246-.

11. Kalsbeek W, Heiss G. Building bridges between populations and samples in epidemiological studies. Annual Review of Public Health. 2000;21(1):147-69.

12. Boardman H, Thomas E, Ogden H, Croft P, Millson D. A method to determine if consenters to population surveys are representative of the target study population. Journal of Public Health. 2005;27(2):212.

13. Dekkers OM, von Elm E, Algra A, Romijn JA, Vandenbroucke JP. How to assess the external validity of therapeutic trials: a conceptual approach. International Journal of Epidemiology. 2010;39(1):89-94.

14. Torner A, Duberg AS, Dickman P, Svensson A. A Proposed Method to Adjust for Selection Bias in Cohort Studies. American Journal of Epidemiology. 2010;171(5):602-8.

15. Jonsson S, Hedblad B, Engstrom G, Nilsson P, Berglund G, Janzon I. Influence of obesity on cardiovascular risk. Twenty-three-year follow-up of 22 025 men from an urban Swedish population. International Journal of Obesity. 2002;26(8):1046-53.

16. Brewster D, Stockton D, Harvey J, Mackay M. Reliability of cancer registration data in Scotland, 1997. European Journal of Cancer. 2002;38(3):414-7.

17. Molster C, Bower C, O'Leary P. Community attitudes to the collection and use of identifiable data for health research–is it an invasion of privacy? Australian and New Zealand Journal of Public Health. 2007;31(4):313-7.

18. Gershon A, Tu J. The effect of privacy legislation on observational research. Canadian Medical Association Journal. 2008;178(7):871.

19. Rothman K. The rise and fall of epidemiology, 1950 2000 AD. New England Journal of Medicine. 1981;304(10):600-2.

20. Australian Institute of Health and Welfare. Diabetes prevalence in Australia: an assessment of national data coverage. Diabetes series no14. Canberra: AIHW; 2009.

21. Australian Bureau of Statistics. Census of Population and Housing: Socio-Economic Indexes for Area's (SEIFA). Canberra: ABS; 2004.

22. Alonso A, Seguí-Gómez M, de Irala J, Sánchez-Villegas A, Beunza J, Martínez-Gonzalez M. Predictors of follow-up and assessment of selection bias from dropouts using inverse probability weighting in a cohort of university graduates. European journal of epidemiology. 2006;21(5):351-8.

23. Coley N, Gardette V, Toulza O, Gillette-Guyonnet S, Cantet C, Nourhashemi F, et al. Predictive factors of attrition in a cohort of Alzheimer disease patients. Neuroepidemiology. 2008;31(2):69-79.

24. Holden L, Ware R, Passey M. Characteristics of nonparticipants differed based on reason for nonparticipation: a study involving the chronically ill. Journal of clinical epidemiology. 2008;61(7):728-32.

25. Galea S, Tracy M. Participation rates in epidemiologic studies. Annals of epidemiology. 2007;17(9):643-53.

26. Odierna D, Schmidt L. The Effects of Failing to Include Hard-to-Reach Respondents in Longitudinal Surveys. American journal of public health. 2009;99(8):1515.

27. Hazell M, Morris J, Linehan M, Frank P, Frank T. Factors influencing the response to postal questionnaire surveys about respiratory symptoms. Prim Care Respir J. 2009;18(3):165-70.

28. Ware R, Williams G, Aird R. Participants who left a multiple-wave cohort study had similar baseline characteristics to participants who returned. Annals of epidemiology. 2006;16(11):820-3.

29. Brilleman S, Pachana N, Dobson A. The impact of attrition on the representativeness of cohort studies of older people. BMC Medical Research Methodology. 2010;10(1):71.

30. Watson N, Wooden M. Identifying factors affecting longitudinal survey response. In: Lynn P, editor. Methodology of Longitudinal Surveys. New York: Wiley; 2009. p. 157-81.

31. Rojas NL, Sherrit L, Harris S, Knight JR. The Role of Parental Consent in Adolescent Substance Use Research. Journal of Adolescent Health. 2008;42(2):192-7.

32. Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. Written informed consent and selection bias in observational studies using medical records: systematic review. British Medical Journal. 2009;338.

33. Joshy G, Simmons D. Diabetes information systems: A rapidly emerging support for diabetes surveillance and care. Diabetes Technology & Therapeutics. 2006;8(5):587-97.

34. Nickelsen T. Data validity and coverage in the Danish National Health Registry. A literature review. Ugeskrift for laeger. 2001;164(1):33.

35. Choi B, Pak A. Understanding and minimizing epidemiologic bias in public health research. Revue Canadienne De Sante Publique. 2005;96(4).

36. Biemer P, Lyberg L, Wiley J. Introduction to survey quality: Wiley Online Library; 2003.

37. Blumberg S, Luke J. Coverage bias in traditional telephone surveys of low-income and young adults. Public Opinion Quarterly. 2007;71(5):734.

38. Siddiqui O, Flay B, Hu F. Factors affecting attrition in a longitudinal smoking prevention study. Preventive Medicine. 1996;25(5):554-60.

39. Che YH, Assanangkornchai S, McNeil E, Chongsuvivatwong V, Li JH, Geater A, et al. Predictors of early dropout in methadone maintenance treatment program in Yunnan province, China. Drug and Alcohol Review. 2010;29(3):263-70.

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

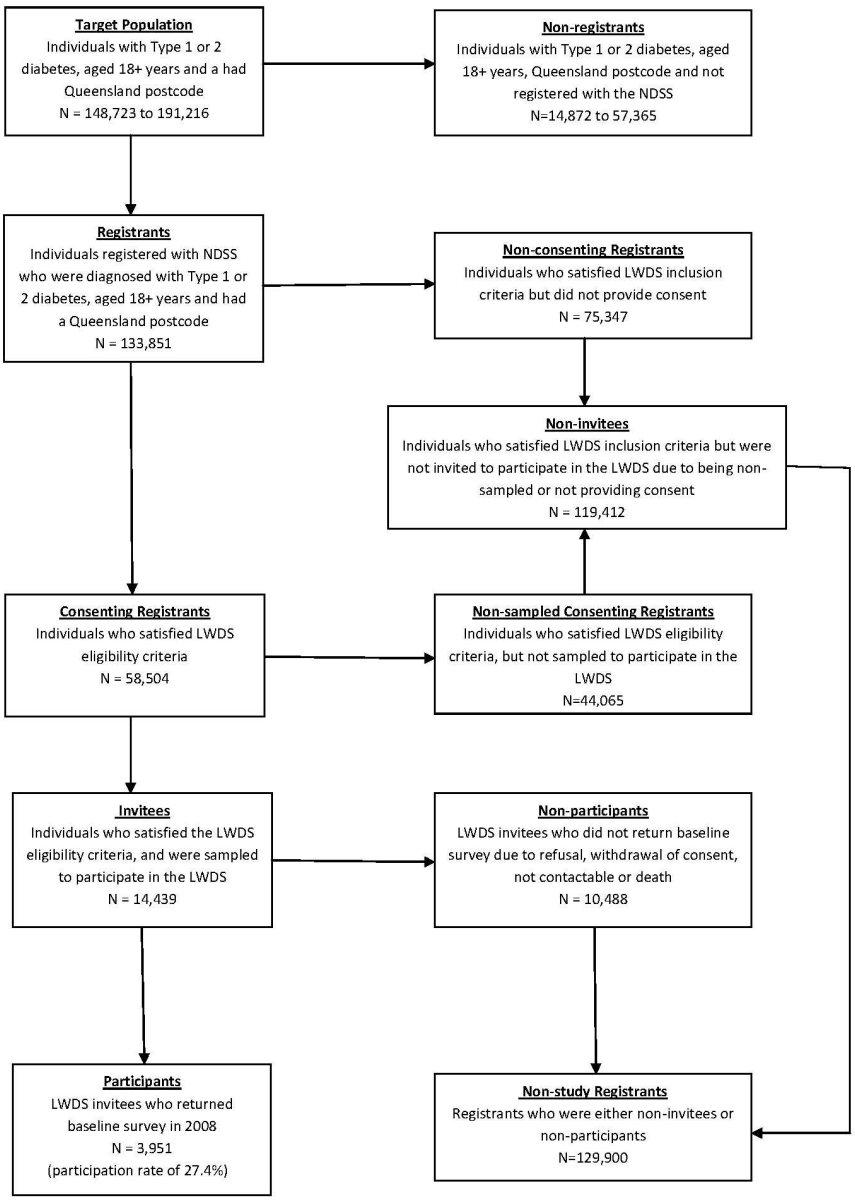
	No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **(See pages 1, 2 & 3 of manuscript)**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **(See pages 2 & 3 of manuscript)**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **(See paragraph 1)**
Objectives	3	State specific objectives, including any prespecified hypotheses **(See paragraph 2)**
Methods		
Study design	4	Present key elements of study design early in the paper **(See paragraph 2)**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **(See paragraph 1)**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **(See paragraph 1 & data only collected from baseline survey)**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed **(Not applicable)**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **(See paragraph 2)**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **(See paragraph 1)**
Bias	9	Describe any efforts to address potential sources of bias **(See paragraph 1 i.e. randomly sampled from disease register)**
Study size	10	Explain how the study size was arrived at **(Not reported in this study as outcomes associated with diabetes not an objective of this paper)**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **(See paragraph 3)**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **(Multivariate logistic analysis was used to control for potential confounders i.e. see paragraph 3)**
		(*b*) Describe any methods used to examine subgroups and interactions **(See paragraph 3)**
		(*c*) Explain how missing data were addressed **(Due to study's objectives, there was no need to incorporate or utilize any statistical adjustment for missing data).**
		(*d*) If applicable, explain how loss to follow-up was addressed **(See paragraph 2, for strategies to maximize participation at baseline)**

		(*e*) Describe any sensitivity analyses **(See paragraph 2 in Results section)**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **(See Figure 1)**
		(b) Give reasons for non-participation at each stage **(See paragraphs 2&3 for associations between baseline characteristics and participation status)**
		(c) Consider use of a flow diagram **(See Figure 1)**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **(See Table 1)**
		(b) Indicate number of participants with missing data for each variable of interest **(See Table 1)**
		(c) Summarise follow-up time (eg, average and total amount) **(Not applicable as only participation at baseline/wave 1 considered)**
Outcome data	15*	Report numbers of outcome events or summary measures over time **(See Table 1)**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **(See Table 1 and paragraph 2)**
		(*b*) Report category boundaries when continuous variables were categorized **(See paragraph 3 in Methods and Table 1)**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **(Not applicable)**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **(See paragraph 2)**
Discussion		
Key results	18	Summarise key results with reference to study objectives **(See paragraph 1, 2 & 3)**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **(See paragraph 4)**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **(See paragraphs 2,3 & 4)**
Generalisability	21	Discuss the generalisability (external validity) of the study results **(See paragraph 4)**
Other information		
	22	Give the source of funding and the role of the funders for the present study and, if

Funding	applicable, for the original study on which the present article is based **(See funding statement in manuscript)**

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.



Participation flowchart
174x241mm (200 x 200 DPI)

BMJ open

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

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2011-000078.R1
Article Type:	Research
Date Submitted by the Author:	11-May-2011
Complete List of Authors:	David, Michael; The University of Queensland, School of Population Health Ware, Robert; The University of Queensland, School of Population Health Donald, Maria; The University of Queensland, School of Population Health Alati, Rosa; The University of Queensland, School of Population Health
Subject Heading:	Statistics & research methods
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS, GENERAL DIABETES < DIABETES & ENDOCRINOLOGY

TITLE PAGE

TITLE:

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

CORRESPONDING AUTHOR:

Name: Michael David
Postal Address: School of Population Health, The University of Queensland, Herston, Queensland 4000, Australia
E-mail: michael.david@uqconnect.edu.au
Tel: +61 733655298
Fax: +61 733655298

CO-AUTHORS:

Robert Ware[1]; Maria Donald[1]; and Rosa Alati[1]
[1] School of Population Health, The University of Queensland, Herston, Queensland, Australia

KEY WORDS and PHRASES:

Generalizability; Cohort Study; Research Consent; Disease Register

WORD COUNT:

2,175

ABSTRACT

Objectives

The knowledge of a study population's similarity to the target population allows researchers to assess the generalizability of their results. Often generalizability 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 3,951 study participants with 10,488 non-participants (individuals who were invited to participate but declined), and with 129,900 non-study registrants (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 to 69, have Type 2 diabetes not requiring insulin, be recently registered and be non-indigenous Australians. Study participants were more likely than non-participants to be aged 50 to 69, have Type 1 diabetes, and be non-indigenous Australians.

Conclusions

The interpretation of a study's generalizability can alter depending on which non-participating group is compared with participants. When assessing generalizability, 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.

Article Summary

Article Focus

* to assess the similarity between participants recruited into a study who were sampled from a chronic disease register, with registrants who did not participate in the study
* to assess whether the differences between participants and registrants who did not participate are similar to the differences between participants and individuals who were invited to participate but declined

Key Messages

* the generalizability of a study should be assessed by comparing participants 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

Strengths and Limitations

* information is available for all individuals registered with the chronic disease register
* the chronic disease register from which study participants were recruited from has high coverage of target population
* only aggregated data was available for registrants who were not invited to participate in the study

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.[1, 2] However the utility of each study's findings depends on whether the results are sufficiently generalizable to the population under investigation. The extent of a study's generalizability, or external validity, depends on how representative of the target population the study's participants are.[3-5] Since information on the target population is often unavailable, investigations concerning the generalizability 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.[6-9] However, a high degree of similarity between participants and non-participants does not necessarily mean results arising from the study will have good generalizability, as these two groups as a whole might not fully represent the target population due to coverage error.[10, 11] The extent to which findings are generalizable 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.[12, 13] The characteristics of this larger group can be accessed through databases such as national chronic disease registers.[6]

Chronic disease registries are increasingly used to recruit participants to cohort studies. One purported advantage of this is to ensure generalizability to the target population.[14-16] In recent years however, legislative reform concerning privacy issues has been introduced in many countries, including Australia,[17] 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 epidemiologic research.[18, 19] This study investigates the generalizability 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 named Diabetes Australia. The NDSS's coverage of Queenslanders with diabetes is estimated to be between 80% and 90%.[20] Since

2001, individuals joining the NDSS have been asked whether they would like to be informed about opportunities to participate in research. Those who consented to be contacted for research purposes and had a valid postal address were 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, in addition to approximately one in six eligible individuals from the rest of Queensland.**

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 maximize participation included reminder cards, telephone calls and replacement surveys. We categorized 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. Individuals in either of these groups were defined as invitees. 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 participate in the LWDS. Individuals in either of these two latter groups were defined as non-invitees. Initially, participants were compared to non-participants (the reference group). For a secondary comparative analysis, the reference group was expanded by the inclusion of non-invitees. Those in this expanded reference group were defined as non-study registrants. Figure 1 depicts the relationships that exist between these groups by way of a schematic diagram. Available individual-level information on participants and non-participants consisted of sex, age, diabetes status, year of NDSS registration, postcode and indigenous status. Postcodes were matched to the Australian Bureau of Statistics' Index of Relative Socio-economic Disadvantage (SEIFA) ranking, and categorised into tertiles.[21] Due to privacy and research consent issues, only covariate aggregate data was available for non-invitees. Ethics approval was obtained from the University of Queensland's Behavioural and Social Sciences Ethical Review Committee.

[Figure 1 to be inserted here]

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 and 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 multivariate 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 was 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 odds ratios (ORs) and 95% confidence intervals (95% CIs).

RESULTS

At the 30th of 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 research and were excluded **(Figure 2)**. Of the remaining 58,504 registrants, 14,439 were invited to participate in the LWDS**;** 3,951 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, 56,264 registrations and 1,711 postcodes were not available for the analyses.

[Figure 2 to be inserted here]

Table 1 displays a comparison of 3,951 participants and 10,488 non-participants, and a comparison between participants and 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**-0.71) or older (0.89; 0.81-0.99)** than those aged 50 to 69 years; and had identified themselves as being indigenous Australians (0.61; **0.48-0.77)**. Those who had Type 1 diabetes **(1.50; 1.19-1.90)** 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 except age **(not shown in Table 1)**, the odds of being a participant if an individual had Type 1 diabetes was **1.13 (0.91 –1.42)** greater than if an individual had Type 2 diabetes and not insulin requiring, while for the fully adjusted model it was **1.50 (1.19–1.90).**

The comparative analyses between participants and non-study registrants (Table 1) shows a number of associational differences when compared to the previous multivariate analysis. The most noticeable is the relationship between participation and diabetes status, as it varies not only in strength, but direction. These analyses

Table 1: **Comparison between Living with Diabetes Participants, Non-participants and Non-study registrants.**

	Participants	Non-participants	Non-study Registrants	Participants Vs Non-participants		Participants Vs Non-study Registrants
	N = 3,951	N = 10,488	N = 129,900	Crude OR (95% CI)	Adjusted OR (95% CI)*	Crude OR (95% CI)
Sex						
Male	2,176 (55.1%)	5,885 (56.1%)	68,618 (52.8%)	**1.00**	**1.00**	1.00
Female	1,775 (44.9%)	4,603 (43.9%)	61,282 (47.2%)	**1.01 (0.93 – 1.10)**	**1.07 (0.98 – 1.17)**	0.91 (0.86 – 0.97)
Age						
18-49	618 (15.6%)	2,246 (21.4%)	21,387 (16.5%)	**0.67 (0.60 - 0.75)**	**0.63 (0.55 – 0.71)**	0.71 (0.66 – 0.79)
50 - 69	2,375 (60.1%)	5,649 (53.9%)	58,988 (45.4%)	**1.00**	**1.00**	1.00
70+	958 (24.3%)	2,593 (24.7%)	49,525 (38.1%)	**0.88 (0.80 – 0.97)**	**0.89 (0.80 - 0.99)**	0.48 (0.45 – 0.52)
Diabetes Status						
Type 2 , No Insulin	3,023 (76.5%)	8,024 (76.5%)	82,717 (63.7%)	**1.00**	**1.00**	1.00
Type 2, Insulin	738 (18.7%)	1,986 (18.9%)	28,336 (21.8%)	**0.97 (0.87 – 1.08)**	**0.97 (0.86 – 1.11)**	0.71 (0.66 – 0.77)
Type 1, Insulin	190 (4.8%)	478 (4.6%)	18,847 (14.5%)	**1.11 (0.91 – 1.34)**	**1.50 (1.19 – 1.90)**	0.28 (0.24 – 0.32)
Registration Year						
2001 - 2003	1,303 (38.0%)	3,422 (37.0%)	28,741 (38.8%)	**1.00**	**1.00**	1.00
2004 - 2005	805 (23.4%)	2,239 (24.2%)	20,024 (27.0%)	**0.96 (0.85 – 1.07)**	**0.97 (0.87 – 1.09)**	0.89 (0.81 – 0.97)
2006 - 2008	1,325 (38.6%)	3,580 (38.8%)	25,389 (34.2%)	**1.01 (0.92 – 1.12)**	**1.07 (0.97 – 1.19)**	1.15 (1.06 – 1.24)
SEIFA						
Low	830 (21.0%)	2,491 (23.8%)	27,049 (21.1%)	**1.00**	**1.00**	1.00
Middle	1,543 (39.1%)	3,883 (37.1%)	51,932 (40.5%)	**1.13 (1.01 – 1.27)**	**1.11 (0.98 – 1.25)**	0.97 (0.89 – 1.05)
High	1,572 (39.9%)	4,100 (39.1%)	49,214 (38.4%)	**1.11 (0.99 – 1.24)**	**1.07 (0.95 – 1.21)**	1.04 (0.96 – 1.13)
Indigenous						
No	3,838 (97.2%)	9,969 (95.1%)	124,033 (95.5%)	**1.00**	**1.00**	1.00
Yes	113 (2.8%)	519 (4.9%)	5,867 (4.5%)	**0.57 (0.45 – 0.71)**	**0.61 (0.48 – 0.77)**	0.62 (0.52 – 0.75)

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

show that compared with those with Type 2 diabetes and not insulin dependent, individuals with Type 2 diabetes and insulin dependent (0.71; 0.66-0.77), and Type 1 diabetes (0.28; 0.24-0.32) were less likely to participate. In addition, the association between participation status and sex was also strengthened, with females less likely than males to be participants (0.91; 0.86-0.97). There was no evidence of an association between participation and SEIFA, but this was not the case with Year of NDSS registration, as registration between 2006 and 2008 was positively associated with participation (1.15; 1.06-1.24), whilst registration between 2004 and 2005 was inversely associated (0.89; 0.81-0.97). Associations between participation, and the covariates of age and indigenous status were similar in direction and magnitude with those found by the multivariate analysis, except for those aged at least 70 years, which strengthened inversely (0.48; 0.45-0.52).

DISCUSSION

The differences observed on the comparisons between participants and non-participants, and between participants and non-study registrants, confirm that the extent of a study's generalizability should be established by comparing study participants to a group of individuals which best represents the target population. In this study, those who agreed to participate in the LWDS were significantly different from the non-study registrants over a number of characteristics, with the most notable

being diabetes status. Those with Type 2 diabetes who were insulin requiring, were less likely to participate in the LWDS. Individuals were less likely to be participants if they were insulin requiring, with the odds of participation being 29% less likely for those with Type 2 diabetes who were not 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.[22-24] However, this was not the case when participants 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, though less likely to be invited due to consent issues relating to age of diagnosis,[25] 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.

Though these results are consistent with the literature,[5, 26, 27] 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.[9, 28-30].

Our initial comparative analysis was between participants and non-participants, and relied solely on information from those invited to participate in the study. 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 three in the group of invitees **(4.6%)** when compared to the non-study registrants (**14.5%**). Such underrepresentation is the consequence of Type 1 diabetes being predominately diagnosed during childhood and the NDSS consent protocol,[20] which does not include a systematic updating of consent status at the age of 18 amongst 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 amongst the study participants.[25, 31, 32] 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. Firstly, the generalizability of any study's findings to the target population is very much dependent on register coverage and the quality of its database.[16, 33, 34] Increased levels of coverage and data quality lessen the likelihood of biased sample estimates. [35-37] The coverage of the NDSS is estimated to be between 80% and 90%, which is higher than most diabetes registers,[20, 33] thus giving it the potential to produce sampling frames

of a higher data quality than most. Secondly, in analyses such as these which only utilize one time-point, there is an inability to maximize the information provided by time varying determinants of non-response such as age.[23, 38, 39] Thirdly, due to unavailability of individual-level data for non-invitees from the NDSS, 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 individual 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 generalizability 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 utilized fully in any assessment of generalizability.

ACKNOWLEDGMENTS

We would like to especially thank the participants of the Living with Diabetes Study, without their participation this research would not be possible. Also, our sincere thanks go to Diabetes Australia and the National Diabetes Scheme for working with us to make it possible to recruit participants to the Living with Diabetes Study. We are grateful to Queensland Health and the Australian Research Council for funding of this study. In addition, we would also like to thank all members of the Living with Diabetes Study team for their ongoing support and input.

COMPETING INTERESTS

There are no conflicts of interest with respect to this study as outlined in this paper.

CONTRIBUTORSHIP STATEMENT

I declare that I conceived the study and was the primary author responsible for this final version. In addition, Dr. Robert Ware assisted and advised on conceptualisation, statistical analysis and reviewing, while Dr. Rosa Alati and Dr. Maria Donald assisted in the review and editing of the final version of this paper.

FUNDING

We are grateful to both the Australian Research Council and Queensland Health for providing funding that enabled this study to be commenced at the data collection stage and finalized by the completion of this paper.

REFERENCES

1. Tabák AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimäki M, Witte DR. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. The Lancet. 2009;373(9682):2215-21.
2. Mrena S, Virtanen SM, Laippala P, Kulmala P, Hannila ML, Akerblom HK, et al. Models for predicting type 1 diabetes in siblings of affected children. Diabetes care. 2006;29(3):662-7.
3. Szklo M. Population-based cohort studies. Epidemiologic reviews. 1998;20(1):81.
4. Deeg D. Attrition in longitudinal population studies: Does it affect the generalizability of the findings? An introduction to the series. Journal of clinical epidemiology. 2002;55(3):213-5.
5. Drivsholm T, Eplov L, Davidsen M, Jørgensen T, Ibsen H, Hollnagel H, et al. Representativeness in population-based studies: a detailed description of non-response in a Danish cohort study. Scandinavian journal of public health. 2006;34(6):623-31.
6. Sim J. The external validity of group comparative and single system studies. Physiotherapy. 1995;81(5):263-70.
7. Gerrish K, Lacey A, editors. The Research Process in Nursing. 6th Edition ed. Chichester: John Wiley and Sons; 2010.

8. Barry A. How attrition impacts the internal and external validity of longitudinal research. Journal of School Health. 2005;75(7):4.

9. Livingston PM, Lee SE, McCarty CA, Taylor HR. A comparison of participants with non-participants in a population-based epidemiologic study: The Melbourne Visual Impairment Project. Ophthalmic epidemiology. 1997;42(2):73-81.

10. Groves R, Dillman D, Eltinge J, Little R, Biemer P, Lyberg L, et al. Survey methodology. Technometrics. 2005;47(2):246-.

11. Kalsbeek W, Heiss G. Building bridges between populations and samples in epidemiological studies. Annual Review of Public Health. 2000;21(1):147-69.

12. Boardman H, Thomas E, Ogden H, Croft P, Millson D. A method to determine if consenters to population surveys are representative of the target study population. Journal of Public Health. 2005;27(2):212.

13. Dekkers OM, von Elm E, Algra A, Romijn JA, Vandenbroucke JP. How to assess the external validity of therapeutic trials: a conceptual approach. International Journal of Epidemiology. 2010;39(1):89-94.

14. Torner A, Duberg AS, Dickman P, Svensson A. A Proposed Method to Adjust for Selection Bias in Cohort Studies. American Journal of Epidemiology. 2010;171(5):602-8.

15. Jonsson S, Hedblad B, Engstrom G, Nilsson P, Berglund G, Janzon I. Influence of obesity on cardiovascular risk. Twenty-three-year follow-up of 22 025 men from an urban Swedish population. International Journal of Obesity. 2002;26(8):1046-53.

16. Brewster D, Stockton D, Harvey J, Mackay M. Reliability of cancer registration data in Scotland, 1997. European Journal of Cancer. 2002;38(3):414-7.

17. Molster C, Bower C, O'Leary P. Community attitudes to the collection and use of identifiable data for health research–is it an invasion of privacy? Australian and New Zealand Journal of Public Health. 2007;31(4):313-7.

18. Gershon A, Tu J. The effect of privacy legislation on observational research. Canadian Medical Association Journal. 2008;178(7):871.

19. Rothman K. The rise and fall of epidemiology, 1950 2000 AD. New England Journal of Medicine. 1981;304(10):600-2.

20. Welfare AIoHa. Diabetes prevalence in Australia: an assessment of national data coverage. Diabetes series no14. Canberra: AIHW; 2009.

21. ABS. Census of Population and Housing: Socio-Economic Indexes for Area's (SEIFA). Canberra: ABS; 2004.

22. Alonso A, Seguí-Gómez M, de Irala J, Sánchez-Villegas A, Beunza J, Martínez-Gonzalez M. Predictors of follow-up and assessment of selection bias from dropouts using inverse probability weighting in a cohort of university graduates. European journal of epidemiology. 2006;21(5):351-8.

23. Coley N, Gardette V, Toulza O, Gillette-Guyonnet S, Cantet C, Nourhashemi F, et al. Predictive factors of attrition in a cohort of Alzheimer disease patients. Neuroepidemiology. 2008;31(2):69-79.

24. Holden L, Ware R, Passey M. Characteristics of nonparticipants differed based on reason for nonparticipation: a study involving the chronically ill. Journal of clinical epidemiology. 2008;61(7):728-32.

25. Galea S, Tracy M. Participation rates in epidemiologic studies. Annals of epidemiology. 2007;17(9):643-53.

26. Odierna D, Schmidt L. The Effects of Failing to Include Hard-to-Reach Respondents in Longitudinal Surveys. American journal of public health. 2009;99(8):1515.

27. Hazell M, Morris J, Linehan M, Frank P, Frank T. Factors influencing the response to postal questionnaire surveys about respiratory symptoms. Prim Care Respir J. 2009;18(3):165-70.

28. Ware R, Williams G, Aird R. Participants who left a multiple-wave cohort study had similar baseline characteristics to participants who returned. Annals of epidemiology. 2006;16(11):820-3.

29. Brilleman S, Pachana N, Dobson A. The impact of attrition on the representativeness of cohort studies of older people. BMC Medical Research Methodology. 2010;10(1):71.

30. Watson N, Wooden M. Identifying factors affecting longitudinal survey response. In: Lynn P, editor. Methodology of Longitudinal Surveys. New York: Wiley; 2009. p. 157-81.

31. Rojas NL, Sherrit L, Harris S, Knight JR. The Role of Parental Consent in Adolescent Substance Use Research. Journal of Adolescent Health. 2008;42(2):192-7.

32. Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. Written informed consent and selection bias in observational studies using medical records: systematic review. British Medical Journal. 2009;338.

33. Joshy G, Simmons D. Diabetes information systems: A rapidly emerging support for diabetes surveillance and care. Diabetes Technology & Therapeutics. 2006;8(5):587-97.

34. Nickelsen T. Data validity and coverage in the Danish National Health Registry. A literature review. Ugeskrift for laeger. 2001;164(1):33.

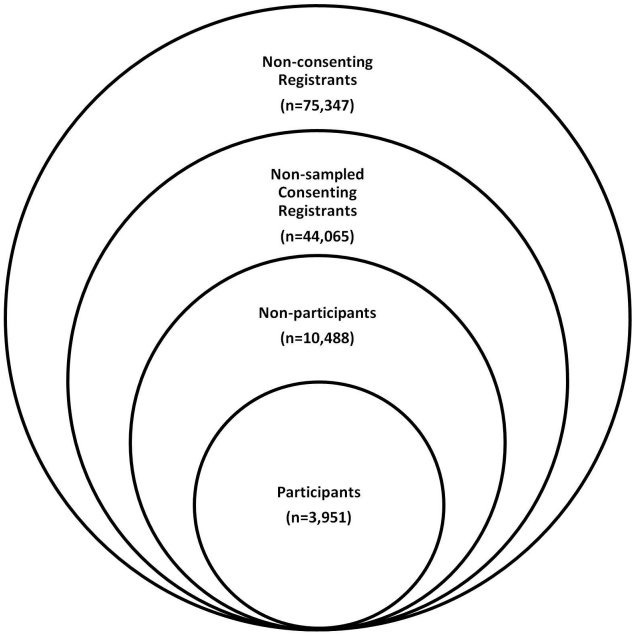
35. Choi B, Pak A. Understanding and minimizing epidemiologic bias in public health research. Revue Canadienne De Sante Publique. 2005;96(4).

36. Biemer P, Lyberg L, Wiley J. Introduction to survey quality. ed. Hoboken, NJ: Wiley Online Library; 2003.

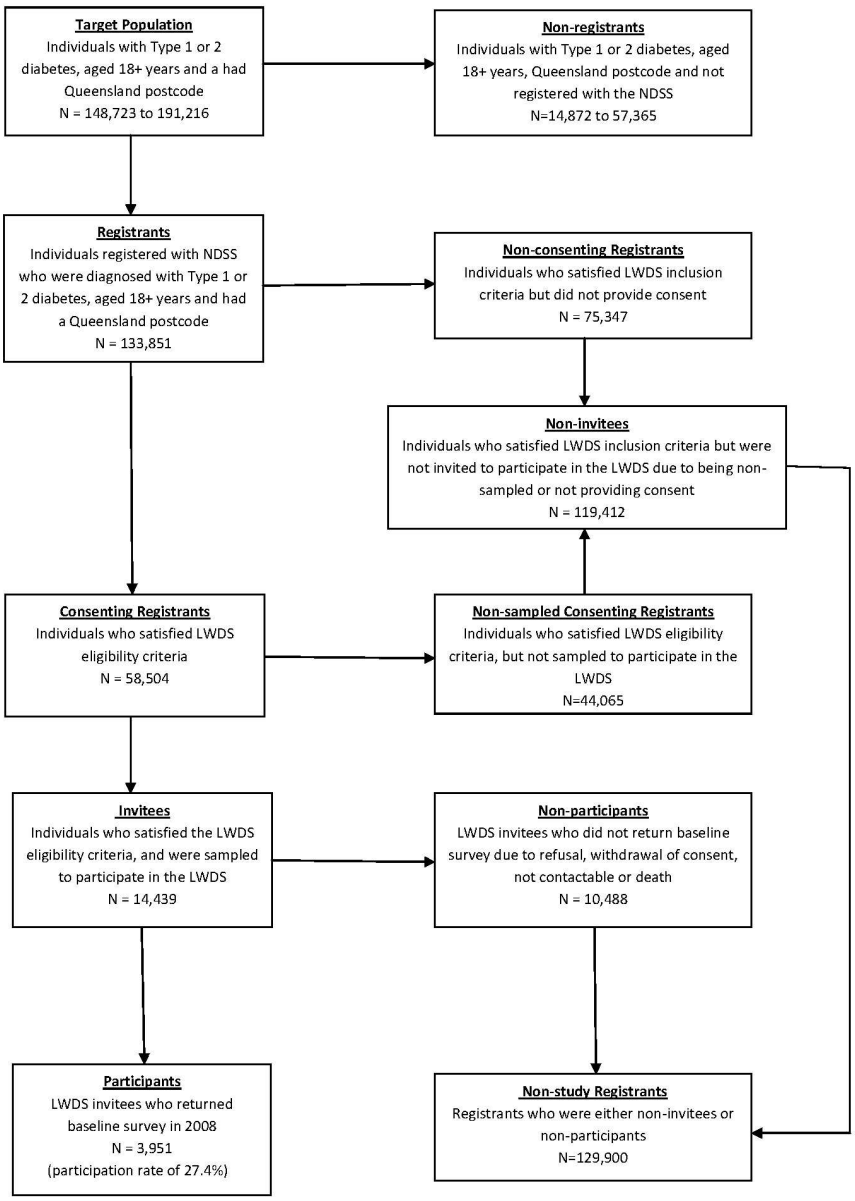
37. Blumberg S, Luke J. Coverage bias in traditional telephone surveys of low-income and young adults. Public Opinion Quarterly. 2007;71(5):734.

38. Siddiqui O, Flay B, Hu F. Factors affecting attrition in a longitudinal smoking prevention study. Preventive Medicine. 1996;25(5):554-60.

39. Che YH, Assanangkornchai S, McNeil E, Chongsuvivatwong V, Li JH, Geater A, et al. Predictors of early dropout in methadone maintenance treatment program in Yunnan province, China. Drug and Alcohol Review. 2010;29(3):263-70.



Schematic diagram of the four mutually exclusive subgroups that constitute registrants with the National Diabetes Scheme.
190x254mm (300 x 300 DPI)



Participation flowchart for the Living with Diabetes Study
174x241mm (200 x 200 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **(See pages 1, 2 & 3 of manuscript)**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **(See pages 2 & 3 of manuscript)**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **(See paragraph 1)**
Objectives	3	State specific objectives, including any prespecified hypotheses **(See paragraph 2)**
Methods		
Study design	4	Present key elements of study design early in the paper **(See paragraph 2)**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **(See paragraph 1)**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **(See paragraph 1 & data only collected from baseline survey)**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed **(Not applicable)**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **(See paragraph 2)**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **(See paragraph 1)**
Bias	9	Describe any efforts to address potential sources of bias **(See paragraph 1 i.e. randomly sampled from disease register)**
Study size	10	Explain how the study size was arrived at **(Not reported in this study as outcomes associated with diabetes not an objective of this paper)**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **(See paragraph 3)**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **(Multivariate logistic analysis was used to control for potential confounders i.e. see paragraph 3)**
		(*b*) Describe any methods used to examine subgroups and interactions **(See paragraph 3)**
		(*c*) Explain how missing data were addressed **(Due to study's objectives, there was no need to incorporate or utilize any statistical adjustment for missing data).**
		(*d*) If applicable, explain how loss to follow-up was addressed **(See paragraph 2, for strategies to maximize participation at baseline)**

		(*e*) Describe any sensitivity analyses **(See paragraph 2 in Results section)**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **(See Figure 1)**
		(b) Give reasons for non-participation at each stage **(See paragraphs 2&3 for associations between baseline characteristics and participation status)**
		(c) Consider use of a flow diagram **(See Figure 1)**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **(See Table 1)**
		(b) Indicate number of participants with missing data for each variable of interest **(See Table 1)**
		(c) Summarise follow-up time (eg, average and total amount) **(Not applicable as only participation at baseline/wave 1 considered)**
Outcome data	15*	Report numbers of outcome events or summary measures over time **(See Table 1)**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **(See Table 1 and paragraph 2)**
		(*b*) Report category boundaries when continuous variables were categorized **(See paragraph 3 in Methods and Table 1)**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **(Not applicable)**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **(See paragraph 2)**
Discussion		
Key results	18	Summarise key results with reference to study objectives **(See paragraph 1, 2 & 3)**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **(See paragraph 4)**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **(See paragraphs 2,3 & 4)**
Generalisability	21	Discuss the generalisability (external validity) of the study results **(See paragraph 4)**
Other information		
	22	Give the source of funding and the role of the funders for the present study and, if

Funding	applicable, for the original study on which the present article is based **(See funding statement in manuscript)**

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

BMJ open

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

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2011-000078.R2
Article Type:	Research
Date Submitted by the Author:	24-Jun-2011
Complete List of Authors:	David, Michael; The University of Queensland, School of Population Health Ware, Robert; The University of Queensland, School of Population Health Donald, Maria; The University of Queensland, School of Population Health Alati, Rosa; The University of Queensland, School of Population Health
Primary Subject Heading:	Statistics & research methods
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS, GENERAL DIABETES < DIABETES & ENDOCRINOLOGY

TITLE PAGE

TITLE:

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

CORRESPONDING AUTHOR:

Name: Michael David
Postal Address: School of Population Health, The University of Queensland, Herston, Queensland 4000, Australia
E-mail: michael.david@uqconnect.edu.au
Tel: +61 733655298
Fax: +61 733655298

CO-AUTHORS:

Robert Ware[1]; Maria Donald[1]; and Rosa Alati[1]

[1] School of Population Health, The University of Queensland, Herston, Queensland, Australia

KEY WORDS and PHRASES:

Generalisability; Cohort Study; Research Consent; Disease Register

WORD COUNT:

2,170

ABSTRACT

Objectives

The 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 3,951 study participants with 10,488 non-participants (individuals who were invited to participate but declined), and with 129,900 non-study registrants (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 to 69, have Type 2 diabetes not requiring insulin, be recently registered and be non-indigenous Australians. Study participants were more likely than non-participants to be aged 50 to 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.

Article Summary

Article Focus

* to assess the similarity between participants recruited into a study who were sampled from a chronic disease register, with registrants who did not participate in the study
* to assess whether the differences between participants and registrants who did not participate are similar to the differences between participants and individuals who were invited to participate but declined

Key Messages

* the generalisability of a study should be assessed by comparing participants 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

Strengths and Limitations

* information is available for all individuals registered with the chronic disease register
* the chronic disease register from which study participants were recruited from has high coverage of target population
* only aggregated data was available for registrants who were not invited to participate in the study

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.(1, 2) 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.(3-5) 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.(6-9) 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.(10, 11) 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.(12, 13) The characteristics of this larger group can be accessed through databases such as national chronic disease registers.(6)

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.(14-16) 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.(17) It is possible that this may seriously limit the usefulness of chronic disease registers for epidemiologic research.(18, 19) 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 named Diabetes Australia. The NDSS's coverage of Queenslanders with diabetes is estimated to be between 80% and 90%. (20) 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). Available individual-level information on participants and non-participants consisted of sex, age, diabetes status, year of NDSS registration, postcode and indigenous status. Postcodes were matched to the Australian Bureau of Statistics' Index of Relative Socio-economic Disadvantage (SEIFA) ranking, and categorised into tertiles.(21) Due to privacy and research consent issues, only covariate aggregate data was available for individuals not invited to participate in the study. Ethics approval was obtained from the University of Queensland's Behavioural and Social Sciences Ethical Review Committee.

[Figure 1 to be inserted here]

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 and 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 multivariate 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 was 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 odds ratios (ORs) and 95% confidence intervals (95% CIs).

RESULTS

At the 30th of 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 research and were excluded (Figure 2). Of the remaining 58,504 registrants, 14,439 were invited to participate in the LWDS; 3,951 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 1,711 postcodes were not available for the analyses.

[Figure 2 to be inserted here]

Table 1 displays a comparison of 3,951 participants and 10,488 non-participants, and a comparison between participants and 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: 0.55-0.71) or older (0.89; 0.81-0.99) than those aged 50 to 69 years; and had identified themselves as being indigenous Australians (0.61; 0.48-0.77). Those who had Type 1 diabetes (1.50; 1.19-1.90) 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 except age (not shown in Table 1), the odds of being a participant if an individual had Type 1 diabetes was 1.13 (0.91 –1.42) greater than if an individual had Type 2 diabetes and was not insulin requiring, while for the fully adjusted model it was 1.50 (1.19–1.90).

The comparative analyses between participants and non-study registrants (Table 1) shows a number of associational differences when compared to the previous multivariate analysis. The most noticeable is the relationship between participation and diabetes status, as it varies not only in strength, but direction. These analyses show that compared to those with Type 2 diabetes who were not insulin reliant, individuals with Type 2 diabetes and insulin reliance (0.71; 0.66-0.77), and Type 1 diabetes (0.28; 0.24-0.32) were less likely to

Table 1: Comparison between Living with Diabetes Participants, Non-participants and Non-study registrants.

	Participants	Non-participants	Non-study Registrants	Participants Vs Non-participants		Participants Vs Non-study Registrants
	N = 3,951	N = 10,488	N = 129,900	Crude OR (95% CI)	Adjusted OR (95% CI)*	Crude OR (95% CI)
Sex						
Male	2,176 (55.1%)	5,885 (56.1%)	68,618 (52.8%)	1.00	1.00	1.00
Female	1,775 (44.9%)	4,603 (43.9%)	61,282 (47.2%)	1.01 (0.93 – 1.10)	1.07 (0.98 – 1.17)	0.91 (0.86 – 0.97)
Age						
18-49	618 (15.6%)	2,246 (21.4%)	21,387 (16.5%)	0.67 (0.60 - 0.75)	0.63 (0.55 – 0.71)	0.71 (0.66 – 0.79)
50 - 69	2,375 (60.1%)	5,649 (53.9%)	58,988 (45.4%)	1.00	1.00	1.00
70+	958 (24.3%)	2,593 (24.7%)	49,525 (38.1%)	0.88 (0.80 – 0.97)	0.89 (0.80 - 0.99)	0.48 (0.45 – 0.52)
Diabetes Status						
Type 2 , No Insulin	3,023 (76.5%)	8,024 (76.5%)	82,717 (63.7%)	1.00	1.00	1.00
Type 2, Insulin	738 (18.7%)	1,986 (18.9%)	28,336 (21.8%)	0.97 (0.87 – 1.08)	0.97 (0.86 – 1.11)	0.71 (0.66 – 0.77)
Type 1, Insulin	190 (4.8%)	478 (4.6%)	18,847 (14.5%)	1.11 (0.91 – 1.34)	1.50 (1.19 – 1.90)	0.28 (0.24 – 0.32)
Registration Year						
2001 - 2003	1,303 (38.0%)	3,422 (37.0%)	28,741 (38.8%)	1.00	1.00	1.00
2004 - 2005	805 (23.4%)	2,239 (24.2%)	20,024 (27.0%)	0.96 (0.85 – 1.07)	0.97 (0.87 – 1.09)	0.89 (0.81 – 0.97)
2006 - 2008	1,325 (38.6%)	3,580 (38.8%)	25,389 (34.2%)	1.01 (0.92 – 1.12)	1.07 (0.97 – 1.19)	1.15 (1.06 – 1.24)
SEIFA						
Low	830 (21.0%)	**2**,491 (23.8%)	27,049 (21.1%)	1.00	1.00	1.00
Middle	1,543 (39.1%)	3,883 (37.1%)	51,932 (40.5%)	1.13 (1.01 – 1.27)	1.11 (0.98 – 1.25)	0.97 (0.89 – 1.05)
High	1,572 (39.9%)	4,100 (39.1%)	49,214 (38.4%)	1.11 (0.99 – 1.24)	1.07 (0.95 – 1.21)	1.04 (0.96 – 1.13)
Indigenous						
No	3,838 (97.2%)	9,969 (95.1%)	124,033 (95.5%)	1.00	1.00	1.00
Yes	113 (2.8%)	519 (4.9%)	5,867 (4.5%)	0.57 (0.45 – 0.71)	0.61 (0.48 – 0.77)	0.62 (0.52 – 0.75)

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

participate. In addition, the association between participation status and sex was also strengthened, with females less likely than males to be participants (0.91; 0.86-0.97). There was no evidence of an association between participation and SEIFA, but this was not the case with Year of NDSS registration, as registration between 2006 and 2008 was positively associated with participation (1.15; 1.06-1.24), whilst registration between 2004 and 2005 was inversely associated (0.89; 0.81-0.97). Associations between participation, and the covariates of age and indigenous status were similar in direction and magnitude with those found by the multivariate analysis, except for those aged at least 70 years, which strengthened inversely (0.48; 0.45-0.52).

DISCUSSION

The differences observed on the comparisons between participants and non-participants, and between participants and non-study registrants, confirm that the extent of a study's generalisability should be established by comparing study participants to a group of individuals which best represents the target population. In this study, those who agreed to participate in the LWDS were significantly different from the non-study registrants over a number of characteristics, with the most notable being diabetes status. Those with Type 2 diabetes who were insulin requiring, were less likely to participate in the LWDS. Individuals were less likely to be participants if they 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.(22-24) However, this was not the case when participants 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, though less likely to be invited due to consent issues relating to age of diagnosis,(25) 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.

Though these results are consistent with the literature,(5, 26, 27) 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.(9, 28-30).

Our initial comparative analysis was between participants and non-participants, and relied solely on information from those invited to participate in the study. 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 three 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,(20) which does not include a systematic updating of consent status at the age of 18 amongst 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 amongst the study participants.(25, 31, 32) 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. Firstly, the generalisability of any study's findings to the target population is very much dependent on register coverage and the quality of its database.(16, 33, 34) Increased levels of coverage and data quality lessen the likelihood of biased sample estimates. (35-37) The coverage of the NDSS is estimated to be between 80% and 90%, which is higher than most diabetes registers,(20, 33) thus giving it the potential to produce sampling frames of a higher data quality than most. Secondly, 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.(23, 38, 39) Thirdly,

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 individual 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.

ACKNOWLEDGMENTS

We would like to especially thank the participants of the Living with Diabetes Study; without their participation this research would not be 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. We are also grateful to Queensland Health and the Australian Research Council for the funding of this study. In addition, we would also like to thank all members of the Living with Diabetes Study team for their ongoing support and input.

COMPETING INTERESTS

There are no conflicts of interest with respect to this study as outlined in this paper.

CONTRIBUTORSHIP STATEMENT

All authors designed the study. M. Donald was responsible for the 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.

FUNDING

This research was funded by Australian Research Council (DP0988805) and Queensland Health through the Queensland Strategy for Chronic Disease 2005 – 20015.

REFERENCES

1. Tabák AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimäki M, Witte DR. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. The Lancet. 2009;373(9682):2215-21.
2. Mrena S, Virtanen SM, Laippala P, Kulmala P, Hannila ML, Akerblom HK, et al. Models for predicting type 1 diabetes in siblings of affected children. Diabetes Care. 2006;29(3):662-7.
3. Szklo M. Population-based cohort studies. Epidemiol Rev. 1998;20(1):81-90.
4. Deeg DJH. Attrition in longitudinal population studies: Does it affect the generalizability of the findings? An introduction to the series. J Clin Epidemiol. 2002;55(3):213-5.
5. Drivsholm T, Eplov LF, Davidsen M, Jørgensen T, Ibsen H, Hollnagel H, et al. Representativeness in population-based studies: a detailed description of non-response in a Danish cohort study. Scand J Public Health. 2006;34(6):623-31.
6. Sim J. The external validity of group comparative and single system studies. Physiotherapy. 1995;81(5):263-70.
7. Gerrish K, Lacey A, editors. The Research Process in Nursing. 6th Edition ed. Chichester: John Wiley and Sons; 2010.
8. Barry AE. How attrition impacts the internal and external validity of longitudinal research. J Sch Health. 2005;75(7):267-70.

9. Livingston PM, Lee SE, McCarty CA, Taylor HR. A comparison of participants with non-participants in a population-based epidemiologic study: The Melbourne Visual Impairment Project. Ophthalmic Epidemiol. 1997;42(2):73-81.

10. Groves RM, Floyd J, Fowler JR, Couper JM, Singer E, Tourangeau R. Survey methodology. Hoboken, NJ: Wiley; 2004.

11. Kalsbeek W, Heiss G. Building bridges between populations and samples in epidemiological studies. Annu Rev Public Health. 2000;21(1):147-69.

12. Boardman HF, Thomas E, Ogden H, Croft PR, Millson DS. A method to determine if consenters to population surveys are representative of the target study population. J Public Health. 2005;27(2):212-4.

13. Dekkers OM, von Elm E, Algra A, Romijn JA, Vandenbroucke JP. How to assess the external validity of therapeutic trials: a conceptual approach. Int J Epidemiol. 2010;39(1):89-94.

14. Torner A, Duberg AS, Dickman P, Svensson A. A Proposed Method to Adjust for Selection Bias in Cohort Studies. Am J Epidemiol. 2010;171(5):602-8.

15. Jonsson S, Hedblad B, Engstrom G, Nilsson P, Berglund G, Janzon I. Influence of obesity on cardiovascular risk. Twenty-three-year follow-up of 22 025 men from an urban Swedish population. Int J Obes. 2002;26(8):1046-53.

16. Brewster DH, Stockton D, Harvey J, Mackay M. Reliability of cancer registration data in Scotland, 1997. Eur J Cancer. 2002;38(3):414-7.

17. Molster C, Bower C, O'Leary P. Community attitudes to the collection and use of identifiable data for health research–is it an invasion of privacy? Aust N Z J Public Health. 2007;31(4):313-7.

18. Gershon AS, Tu JV. Data validity and coverage in the Danish National Health Registry. A literature review. Can Med Assoc J. 2008;178(7):871-3.

19. Rothman KJ. The rise and fall of epidemiology, 1950 2000 AD. N Engl J Med. 1981;304(10):600-2.

20. Australian Institute of Health and Welfare. Diabetes prevalence in Australia: an assessment of national data coverage. Diabetes series no14. Canberra: AIHW; 2009.

21. Australian Bureau of Statistics. Census of Population and Housing: Socio-Economic Indexes for Area's (SEIFA). Canberra: ABS; 2004.

22. Alonso A, Seguí-Gómez M, de Irala J, Sánchez-Villegas A, Beunza J, Martínez-Gonzalez MÁ. Predictors of follow-up and assessment of selection bias from dropouts using inverse probability weighting in a cohort of university graduates. Eur J Epidemiol. 2006;21(5):351-8.

23. Coley N, Gardette V, Toulza O, Gillette-Guyonnet S, Cantet C, Nourhashemi F, et al. Predictive factors of attrition in a cohort of Alzheimer disease patients. Neuroepidemiology. 2008;31(2):69-79.

24. Holden L, Ware RS, Passey M. Characteristics of nonparticipants differed based on reason for nonparticipation: a study involving the chronically ill. J Clin Epidemiol. 2008;61(7):728-32.

25. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol. 2007;17(9):643-53.

26. Odierna DH, Schmidt LA. The Effects of Failing to Include Hard-to-Reach Respondents in Longitudinal Surveys. Am J Public Health. 2009;99(8):1515-21.

27. Hazell ML, Morris JA, Linehan MF, Frank PI, Frank TL. Factors influencing the response to postal questionnaire surveys about respiratory symptoms. Prim Care Respir J. 2009;18(3):165-70.

28. Ware RS, Williams GM, Aird RL. Participants who left a multiple-wave cohort study had similar baseline characteristics to participants who returned. Ann Epidemiol. 2006;16(11):820-3.

29. Brilleman SL, Pachana NA, Dobson AJ. The impact of attrition on the representativeness of cohort studies of older people. BMC Med Res Methodol. 2010;10(1):71-9.

30. Watson N, Wooden M. Identifying factors affecting longitudinal survey response. In: Lynn P, editor. Methodology of Longitudinal Surveys. New York: Wiley; 2009. p. 157-81.

31. Rojas NL, Sherrit L, Harris S, Knight JR. The Role of Parental Consent in Adolescent Substance Use Research. J Adolesc Health. 2008;42(2):192-7.

32. Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. Written informed consent and selection bias in observational studies using medical records: systematic review. Br Med J. 2009;338:b866.

33. Joshy G, Simmons D. Diabetes information systems: A rapidly emerging support for diabetes surveillance and care. Diabetes Technol Ther. 2006;8(5):587-97.

34. Nickelsen TN. Data validity and coverage in the Danish National Health Registry. A literature review. Ugeskr Laeger. 2001;164(1):33-7.

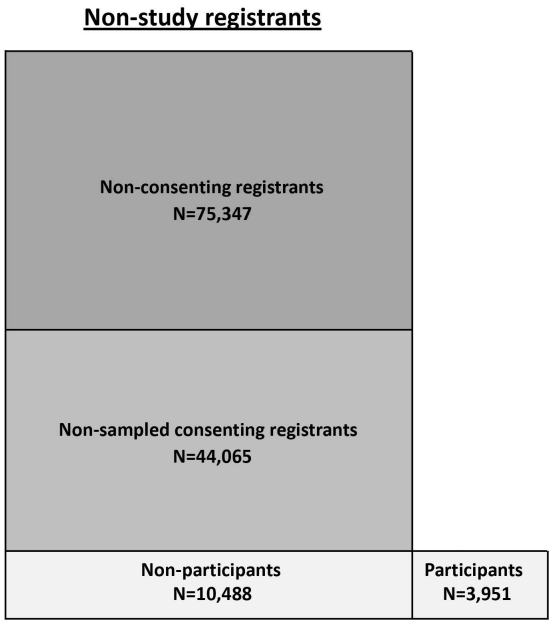
35. Choi BCK, Pak AWP. Understanding and minimizing epidemiologic bias in public health research. Revue Canadienne De Sante Publique. 2005;96:284-6.

36. Biemer PP, Lyberg L, Wiley J. Introduction to survey quality. ed. Hoboken, NJ: Wiley Online Library; 2003.

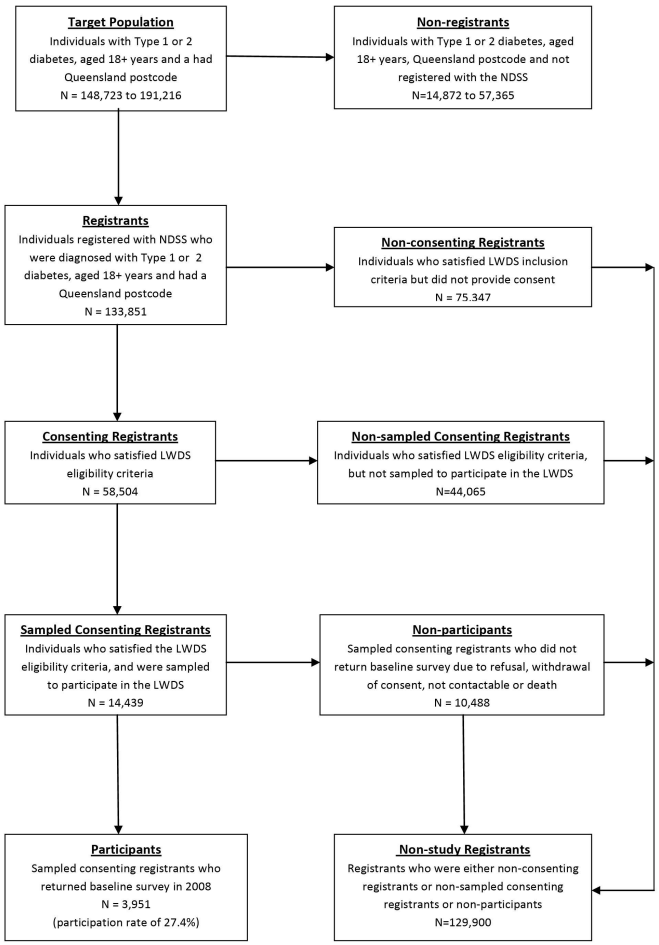
37. Blumberg SJ, Luke JV. Coverage bias in traditional telephone surveys of low-income and young adults. Public Opin Q. 2007;71(5):734-49.

38. Siddiqui O, Flay BR, Hu FB. Factors affecting attrition in a longitudinal smoking prevention study. Prev Med. 1996;25(5):554-60.

39. Che YH, Assanangkornchai S, McNeil E, Chongsuvivatwong V, Li JH, Geater A, et al. Predictors of early dropout in methadone maintenance treatment program in Yunnan province, China. Drug Alcohol Rev. 2010;29(3):263-70.



The four mutually exclusive groups of registrants
190x253mm (300 x 300 DPI)



Participant flowchart in the Living with Diabetes Study
209x296mm (300 x 300 DPI)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract **(See pages 1, 2 & 3 of manuscript)**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found **(See pages 2 & 3 of manuscript)**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **(See paragraph 1)**
Objectives	3	State specific objectives, including any prespecified hypotheses **(See paragraph 2)**
Methods		
Study design	4	Present key elements of study design early in the paper **(See paragraph 2)**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **(See paragraph 1)**
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **(See paragraph 1 & data only collected from baseline survey)**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed **(Not applicable)**
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **(See paragraph 2)**
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **(See paragraph 1)**
Bias	9	Describe any efforts to address potential sources of bias **(See paragraph 1 i.e. randomly sampled from disease register)**
Study size	10	Explain how the study size was arrived at **(Not reported in this study as outcomes associated with diabetes not an objective of this paper)**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **(See paragraph 3)**
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding **(Multivariate logistic analysis was used to control for potential confounders i.e. see paragraph 3)**
		(*b*) Describe any methods used to examine subgroups and interactions **(See paragraph 3)**
		(*c*) Explain how missing data were addressed **(Due to study's objectives, there was no need to incorporate or utilize any statistical adjustment for missing data).**
		(*d*) If applicable, explain how loss to follow-up was addressed **(See paragraph 2, for strategies to maximize participation at baseline)**

		(*e*) Describe any sensitivity analyses **(See paragraph 2 in Results section)**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **(See Figure 1)**
		(b) Give reasons for non-participation at each stage **(See paragraphs 2&3 for associations between baseline characteristics and participation status)**
		(c) Consider use of a flow diagram **(See Figure 1)**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders **(See Table 1)**
		(b) Indicate number of participants with missing data for each variable of interest **(See Table 1)**
		(c) Summarise follow-up time (eg, average and total amount) **(Not applicable as only participation at baseline/wave 1 considered)**
Outcome data	15*	Report numbers of outcome events or summary measures over time **(See Table 1)**
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included **(See Table 1 and paragraph 2)**
		(*b*) Report category boundaries when continuous variables were categorized **(See paragraph 3 in Methods and Table 1)**
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period **(Not applicable)**
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses **(See paragraph 2)**
Discussion		
Key results	18	Summarise key results with reference to study objectives **(See paragraph 1, 2 & 3)**
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias **(See paragraph 4)**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **(See paragraphs 2,3 & 4)**
Generalisability	21	Discuss the generalisability (external validity) of the study results **(See paragraph 4)**
Other information		
	22	Give the source of funding and the role of the funders for the present study and, if

Funding	applicable, for the original study on which the present article is based **(See funding statement in manuscript)**

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.