ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 675−680

Journal of Clinical Epidemiology

Alternative population sampling frames produced important differences in estimates of association: a case−control study of vasculitis

Gary J. Macfarlane[a,*], Gareth T. Jones[a], Leyla Swafe[b], David M. Reid[c], Neil Basu[a]

[a]*Musculoskeletal Research Programme (Epidemiology Group), School of Medicine and Dentistry, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD, UK*

[b]*School of Medicine and Dentistry, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD, UK*

[c]*Musculoskeletal Research Programme, School of Medicine and Dentistry, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD*

Accepted 28 September 2012; Published online 10 January 2013

Abstract

Objective: A common population sampling frame in countries with universal health care is health service registers. We have evaluated the use of such a register, in the United Kingdom, against a commercially available database claiming large population coverage, an alternative that offers ease of access and flexibility of use.

Study Design and Setting: A case−control study of vasculitis, which recruited cases from secondary care clinics in Scotland, compared two alternative sampling frames for population controls, namely the registers of National Health Service (NHS) primary care practices and a commercially available database. The characteristics of controls recruited from both sources were compared in addition to separate case−control comparison using logistic regression.

Results: A total of 166 of 189 cases participated (88% participation rate), while both the commercial database and NHS Central Register (NHSCR) controls achieved a participation rate of 24% among persons assumed to have received the invitation. On several measures, the NHSCR patients reported poorer health than the commercial database controls: low scores on the physical component score of the Short Form 36 (odds ratio [OR]: 2.3; 95% confidence interval [CI]: 1.3−4.1), chronic widespread pain (OR: 2.3; CI: 1.1−4.7), and high levels of fatigue (OR: 2.0; CI: 1.3−3.1). These had an important influence on the estimates of association with case status with one association (pain) showing a strong and significant association using commercial database controls, which was absent with NHSCR controls.

Conclusion: There are important differences in self-reported measures of health and quality of life using controls from two alternative population sampling frames. It emphasizes the importance of methodological rigor and prior assessment in choosing sampling frames for case−control studies.

Keywords: Sampling frame; Participation; Case−control; Odds ratios: health registers; Vasculitis; Bias

1. Introduction

Population-based studies are often considered to be the gold standard in terms of sampling frame for epidemiological studies. They are considered to be less prone to the selection biases that can affect other designs, for example selecting controls in a case−control study from hospital in- or outpatients or through friend controls. Sources of population sampling frames (or approximations to such a sampling frame) differ between countries. In the United Kingdom, an example of a country with universal health care, researchers often use lists of persons registered with local family doctors (general practitioners [GPs]). The National Health Service (NHS) provides universal health care, and it is estimated that 96% of the population are registered with a GP [1] from whom they would receive initial care for most nonemergency medical conditions. It is thus argued that this NHS register provides a convenient population sampling frame for health research. However, a previous study from our own center showed that in Scotland such a register had a coverage of only 85% [2].

There are challenges in continuing to use such a health service sampling frame. First, the legislation aimed at protecting individuals' data, means that the access to data is less easy than earlier. For example, unless there is considered to be a strong public interest, in advance of the individual providing specific consent for the researcher to have access to their contact details, the data has to be processed by a person who is considered to have legitimate access to the data, that is the GP in the case of a practice list.

Conflict of Interest/Financial Disclosure: The authors have no financial disclosures of conflicts of interest to declare.

* Corresponding author.

E-mail address: g.j.macfarlane@abdn.ac.uk (G.J. Macfarlane).

0895-4356/$ - see front matter

http://dx.doi.org/10.1016/j.jclinepi.2012.09.009

What is new?

What is already known?

- It is assumed that sampling frames that provide broad coverage of the general population will result in control populations for case–control studies, which are broadly similar. A commonly used sampling frame in countries with universal health care, such as the United Kingdom, is the health service registration lists.

What this adds to what was known?

- A health service registration list and commercial database with approximately 80% population coverage, when used as sampling frames, produced similar participation rates, populations with different health characteristics, and important differences in the estimates of association. Careful evaluation of population sampling frames needs to be undertaken before assuming their suitability for use.

This is problematic in terms of organizing the initial mailings from participating practices, and the researchers will only receive information on persons who choose to respond. Second, although a researcher is only using the information as a sampling frame for an epidemiological study rather than specifically collecting NHS data, the researcher still needs to apply for ethical approval from an NHS research ethics committee, which has, over recent years, become considerably more resource intensive.

An alternative option is to use a commercially available database. Companies that provide such databases compile information from a variety of sources and access to such information can be purchased. Some claim extensive coverage of the population and provide information on address, age (or age range), and residence history. The information is gathered from publicly available sources. Use of this type of register could be suitable for population-based epidemiological studies and potentially offer a number of advantages in terms of access to data and flexibility of use. However, there has been concern about the use of such data, including persons appearing on registers who may not be representative of the wider population in terms of health or, for example, lifestyle factors that may influence health. In addition, it is conceivable that subjects are less likely to respond to direct contact from an unknown researcher in comparison with a recognized and trusted organization such as the NHS.

We have therefore taken an opportunity afforded by the conduct, across centers in Scotland, of a case–control study of patients with antineutrophil cytoplasmic antibodies (ANCA)-associated vasculitis (AAV) to assess the use of such a commercial database for recruiting controls. AAV is a chronic multisystem autoimmune condition, and one of the aims of this case–control study was to determine the quality of life (QoL) in patients with AAV and specifically assess whether they experienced mood disorders (anxiety and depression), pain, and fatigue more commonly. The study collected data from two alternative population control groups: an NHS register and a commercial population database. We tested the hypotheses that the use of a commercial database as a sampling frame for controls, in comparison with an NHS register, leads to a lower participation rate but no important changes in the assessment of associations of selected health measures with case status.

2. Methods

The "Vasculitis Quality of Life" study used a multicenter case–control design where cases were patients with AAV fulfilling recognized classification criteria [3] and controls were selected from "the general population." All participants completed a questionnaire comprising a number of validated measures of QoL.

2.1. Case and control selection

Cases were recruited from consecutive clinic attendees at recruiting hospitals in Scotland. Willing patients received the study questionnaire, which they were asked to complete at home and return by mail. Two population sampling frames were used to provide separate sets of matched general population controls:

1. Commercial database sampling frame: www.192.com is an online directory that claims to provide the most comprehensive service of its kind in the United Kingdom. Controls were invited within a month of cases completing the study questionnaire. For each case, the search engine of the commercial database provided details of potential controls residing at the same postcode. The resulting lists of records were consecutively searched and those persons matching the relevant case, in terms of sex and age (±5 years), were selected and their addresses noted. Where insufficient controls were identified using the full seven-character postcode, a further list of potential controls was compiled from a broader geographical area applying only the initial four characters of the case's postcode. The first five identified controls were sent a questionnaire. If, after 6 weeks, a case remained unmatched to a commercial database control (i.e., no completed responses from any of the first five invited controls), the remaining identified controls (up to a maximum of five) were sent a questionnaire.
2. NHS Central Register (NHSCR) sampling frame: In Scotland, the NHSCR contains the basic demographic details of all persons currently registered with a GP in