Annals of Epidemiology 28 (2018) 385–391

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

Response rates in case-control studies of cancer by era of fieldwork and by characteristics of study design

Mengting Xu, MSc [a, b, *], Lesley Richardson, MSc [a], Sally Campbell, MSc [a], Javier Pintos, PhD [a], Jack Siemiatycki, PhD [a, b]

[a] University of Montreal Hospital Research Centre (CRCHUM), Montreal, Quebec, Canada
[b] Department of Social and Preventive Medicine, University of Montreal, Montreal, Quebec, Canada

ARTICLE INFO

Article history:
Received 2 August 2017
Accepted 3 April 2018
Available online 9 April 2018

Keywords:
Cancer
Case-control studies
Epidemiologic methods
Participation rate
Response rate

ABSTRACT

Purpose: The purpose of this study was to describe time trends in response rates in case-control studies of cancer and identify study design factors that influence response rate.
Methods: We reviewed 370 case-control studies of cancer published in 12 journals during indicator years in each of the last four decades. We estimated time trends of response rates and reasons for nonresponse in each of the following types of study subjects: cases, medical source controls, and population controls. We also estimated response rates according to characteristics of study context.
Results: Median response rates among cases and population controls were between 75% and 80% in the 1970s. Between 1971 and 2010, study response rates declined by 0.31% per year for cases and 0.78% for population controls. Only a minority of studies reported reasons for nonparticipation; subject refusal was the most common reported reason. Studies conducted in North America had lower median response rates than studies conducted in Europe. In-person and telephone interviews elicited higher response rates than mail questionnaires.
Conclusions: Response rates from case-control studies of cancer have declined, and this could threaten the validity of results derived from these studies.

The case-control study design is often used to examine the etiology of cancer [1,2]. To ensure internal validity, the investigator should aim to enroll representative samples of cases and of controls from the same source population. The selected controls should provide an unbiased estimate of exposure prevalence in the source population that gives rise to the cases, conditional on covariates [3,4]. Selection bias occurs when this principle is violated. One form of such bias is nonresponse bias, which arises when participation is differential by exposure and by disease status [4]. Although studies with low subject response rates do not necessarily produce biased risk estimates, they are more susceptible to nonresponse bias than studies with higher participation.

Time trends in response rates in case-control studies have been reviewed up to the 1990s or early 2000s [5–8]. Two of those reviews [6,7] concluded that there was no evidence of declining response rates up to the late 1990s. The two other reviews [5,8] concluded that there was a decline of response rates up to the early 2000s. These reviews were either narrative or included rather few studies; the largest surveyed 82 case-control studies. They accepted authors' statements about response rates, whereas there is some inconsistency in the way different authors define the denominator for computing response rates. There are clearly different issues in response potential from cases and from controls, but there are also different issues in response potential between controls selected in the general population and controls selected from clients of a health facility (e.g., hospital controls); however, this has rarely been examined in previous reviews. Finally, from the perspective of the mid-2010s, the available information was dated, and we hypothesized that there may have been important changes in response rates in the past decade.

The authors have no conflicts of interest to disclose.
All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

* Corresponding author. University of Montreal Hospital Research Centre (CRCHUM), Tour Saint-Antoine, 850 rue Saint-Denis 3rd Floor, Room S03-614 Montreal, Quebec, H2X 0A9, Canada. Tel.: 514-890-8000ext.31478.
E-mail address: mengting.xu1@gmail.com (M. Xu).

https://doi.org/10.1016/j.annepidem.2018.04.001
1047-2797/

We therefore decided to conduct an assessment of trends in response rates from the 1970s to the 2010s, using standardized definitions of the denominators, and separately for each of three series of subjects, cases, population controls, and medical source controls. We took advantage of this survey to also examine response rates as a function of certain characteristics of the study design or study population. Because of our interest in cancer and because the case-control design is a staple of cancer epidemiology, we focused on response rates in case-control studies of cancer.

Methods

Sample selection

This is a review of questionnaire-based case-control studies of cancer that were published over the past four decades. In a preliminary exercise, we established that a PubMed keyword search and other automatic search methods were not reliable in identifying all case-control studies, and even less, in identifying those that reported response rates. We realized that we would have to review all articles in certain journals one by one. Given the practical limitation to reviewing all journals in all publication years, we instituted a strategy to restrict numbers but yet maintain relevance. Based on our research group's large bibliographic databank of cancer case-control studies published from 1980 to 2014, we identified 12 international journals that seemed to be the main publication sources of relevant case-control studies of cancer during this period. Supplementary Figure 1 shows the list of surveyed journals selected for this review; some of which did not exist for the entire period of observation. For reasons of feasibility, we further restricted attention to articles published in certain calendar years in each decade, namely 1984–1986, 1995, 2005, and 2013. We started this project in 2014; thus, 2013 was chosen as the approximate "mid-decade" year for the 2010s. A 3-year period was selected to represent the mid-1980s because of the small number of studies per year before 1990. For those selected journals and those years, we "manually" examined each article in each issue and selected those publications that satisfied the following additional inclusion criteria:

- Focus on cancer etiology in adults.
- Conducted in North America, Europe, or Australia.
- Used the classic case-control design; nested case-control or case-cohort studies were excluded.
- Included at least 50 cases or controls.
- Data collection from subjects or their proxy respondents using questionnaire instruments.
- If multiple publications were produced based on the same case and control series, only the latest publication was included. We also sought relevant information on subject participation from previous published reports of the study team if such information was missing in the selected publication.

A total of 370 studies providing 370 case series and 422 control series were identified and included in this review.

Data collection

We reviewed all eligible studies and extracted the following information: journal name; publication year; fieldwork period; study population; examined cancer type (grouped by combination of anatomical proximity and usual survival); types of control series (population, medical source, and friends and family control series); mode of data collection (in-person, mail, telephone, or multiple methods); and types of respondent accepted (self only, proxy only, or self and proxy). For each reported subject series, we extracted the frequency distribution of subject participation, including reasons for nonparticipation. Typical reasons for nonparticipation were subject refusal, subject deceased or too ill, subject unreachable, and when applicable, subject not contacted due to medical source obstacles (i.e., physicians refusing access to their patients or medical staff failing to contact the patients).

Response rate calculation

Exclusion of eligible nonrespondents for any reason, whether among cases or controls, could lead to biased risk estimates if participation differs by exposure and by disease status. We defined response rate by the following formulas:

$$\textit{Response Rate} = \frac{\textit{Participants}}{\textit{Eligible Subjects}} \tag{1}$$

$$\textit{Eligible Subjects} = \textit{Participants} + \textit{Subject Refusal} + \textit{Subject Deceased or Too Ill} + \textit{Subject Unreachable} + \textit{Medical Source Obstacles (if applicable)} \tag{2}$$

When the published article provided the number of subjects by each potential outcome category implied by formula [2], we calculated response rates and used those in this article. When studies did not provide sufficient information to allow for our calculation of their response rates, we recorded the rates reported by the authors. In fact, most studies did not provide the requisite information and we usually had to defer to the author's reported response rate.

Data analyses

We reviewed and, if necessary, revised, in each study, the number of subjects who were eligible, who participated, and who failed to participate for each possible reason. We then computed response rates and nonresponse rates by reason for nonparticipation. There were many studies that did not report response rates and many more that did not report one or more of the possible reasons for nonparticipation. This could mean that there were no instances of such nonresponse or it could mean that the authors simply ignored such nonresponse and assumed that those subjects should not be included in the denominator of eligibles. For the purpose of the present analysis, we kept track of the nonmentions and did not assume that these necessarily indicated that there really were no such instances.

For each type of series (cases, controls), we described the distributions of response and nonresponse rates, among those studies that reported such information or that reported information we could use to deduce such rates. Such descriptive analyses were further stratified by subcategories of some characteristics of study