

In an occupational health surveillance study, auxiliary data from administrative health and occupational databases effectively corrected for nonresponse

Gaëlle Santin^{a,*}, Béatrice Geoffroy^a, Laetitia Bénézet^a, Pauline Delézire^a, Juliette Chatelot^a, Rémi Sitta^b, Jean Bouyer^c, Alice Gueguen^b, SNIIR-AM Cohorts Group^{a,b}

^aDepartment of Occupational Health, InVS French Institute for Public Health Surveillance, 12 rue du Val d'Osne, F-94415 Saint-Maurice, France

^bCESP Center for Research in Epidemiology and Population Health U1018—Population-Based Cohorts' Research Platform, INSERM, Hôpital Paul Brousse, 16 Avenue Paul Vaillant-Couturier, F-94807 Villejuif, France

^cCESP Center for Research in Epidemiology and Population Health U1018—Reproduction, Child Development Research Platform, INSERM, Hôpital Kremlin Bicêtre, 82 rue du Général Leclerc F-94276 Le Kremlin-Bicêtre, France

Accepted 25 October 2013; Published online 31 January 2014

Abstract

Objectives: To show how reweighting can correct for unit nonresponse bias in an occupational health surveillance survey by using data from administrative databases in addition to classic sociodemographic data.

Study Design and Setting: In 2010, about 10,000 workers covered by a French health insurance fund were randomly selected and were sent a postal questionnaire. Simultaneously, auxiliary data from routine health insurance and occupational databases were collected for all these workers. To model the probability of response to the questionnaire, logistic regressions were performed with these auxiliary data to compute weights for correcting unit nonresponse. Corrected prevalences of questionnaire variables were estimated under several assumptions regarding the missing data process. The impact of reweighting was evaluated by a sensitivity analysis.

Results: Respondents had more reimbursement claims for medical services than nonrespondents but fewer reimbursements for medical prescriptions or hospitalizations. Salaried workers, workers in service companies, or who had held their job longer than 6 months were more likely to respond. Corrected prevalences after reweighting were slightly different from crude prevalences for some variables but meaningfully different for others.

Conclusion: Linking health insurance and occupational data effectively corrects for nonresponse bias using reweighting techniques. Sociodemographic variables may be not sufficient to correct for nonresponse. © 2014 Elsevier Inc. All rights reserved.

Keywords: Unit nonresponse; Selection bias; Reweighting; Health insurance data; Occupational data; Surveillance

1. Introduction

A decline in participation rates in epidemiological studies has been observed in recent decades [1]. It is a particular concern in epidemiological surveillance surveys that aim to provide descriptive statistics that may be extrapolated to a target population. A nonresponse bias occurs when the response probability (also called response propensity) and the outcome variable are correlated [2]. It can be corrected when this correlation may be completely explained by a known set of variables. Two main techniques can be used for dealing with nonresponse [3]. The first is

imputation, which consists of modeling the outcome variable and replacing each missing item of data by its predicted value. The second is reweighting, which broadly consists of modeling the response probability and then reweighting data by the inverse of the estimated response probability for each subject so-called inverse probability weighting (IPW). The use of imputation is generally recommended for partial nonresponse (subjects answered a questionnaire but did not fill in all the questions), and reweighting is recommended for unit nonresponse (subjects did not answer a questionnaire at all) [3,4]. As we focus on participation in epidemiological studies, we are specifically interested in unit nonresponse and thus in reweighting. Still, it should be noted that imputation, as well as reweighting, require that some data should be known on both respondents and nonrespondents. This may be particularly challenging for unit nonresponse, but it can be done

Conflict of interest: None.

Funding: None.

* Corresponding author. Tel.: +33-(0)155-12-54-28; fax: +33-(0)141-79-67-88.

E-mail address: g.santin@invs.sante.fr (G. Santin).

What is new?**What this adds to what was known?**

- This study shows not only the interest of linking routine health insurance and occupational data to study nonresponse bias but also how these data can be taken into account to use response probability to estimate prevalences by reweighting techniques.

What is the implication and what should change now?

- In an epidemiological surveillance survey, it is not sufficient to correct nonresponse bias solely with sociodemographic variables. The health and occupation-related data available for both respondents and nonrespondents should also be used.

when survey data can be linked to existing databases such as medical administrative databases containing health-care, occupational, or sociodemographic information [5]. The aim is then to model accurately the probability of response using the variables available. Several epidemiological studies have already addressed this issue and have shown that nonresponse is associated with gender, age, marital status, unhealthy lifestyle, healthcare reimbursement, or occupational status [6–10]. Few studies, however, have used these results to correct the prevalence estimates for nonresponse bias [11,12]. Reweighting methods are in fact rarely used and are poorly known in the epidemiological community.

The principal objective of the present study was to show how reweighting can correct for unit nonresponse bias in an occupational surveillance survey by using data from administrative databases related to health and occupation, in addition to the sociodemographic data traditionally used. We then evaluated the impact on prevalence estimates of reweighting corrections with these auxiliary data.

2. Population and methods

2.1. The Coset-MSA cohort

The Coset-MSA study is part of the overall Coset program (Cohort for Epidemiological Surveillance in Connection with Occupation), which aims to study health characteristics and morbidity trends in relation to occupational factors in the French working population [13]. This program relies on data from three cohorts of individuals insured through the three main social welfare funds in France, which cover 95% of the population: the Constances cohort [14], conducted by the French National Institute for Health and Medical Research (INSERM), and the Coset-

MSA and the Coset-RSI cohorts conducted by the French National Institute for Health Surveillance (InVS).

The Coset-MSA cohort focuses on workers in agriculture and related occupations covered by the corresponding insurance fund, the Mutualité Sociale Agricole (MSA). It includes nonsalaried workers (such as farmers and stud farm managers) and salaried workers (farm workers and some bank, insurance, or agricultural cooperative employees). Before setting up the Coset-MSA cohort in the whole of France, a pilot study was conducted in 2010 and was used as a basis for the present report.

The pilot study included workers aged between 18 and 65 years on December 31, 2008, who had worked at least 90 days in a workplace affiliated to the MSA insurance fund in 2008, in one of five French administrative areas (Bouches-du-Rhône, Pas-de-Calais, Pyrénées-Atlantiques, Saône-et-Loire, and Finistère). In each area, 2,000 individuals were randomly selected from the MSA database after stratification for gender, age, and employment status (salaried vs. nonsalaried worker).

The study protocol was approved by the French Institutional Review Committee (CNIL number 909091 and DR-2010-321).

Finally, after excluding persons who could not be contacted by post ($n = 406$) and refusals ($n = 236$) of data extraction from the SNIIR-AM and MSA databases, a total of 9,358 persons were included in the present study.

2.2. Data

Two types of data were collected and matched (only 0.2% of linkages failed because of the change in personal data in the interval between the sampling procedure and database extraction):

1. First, a self-completed postal questionnaire (40 pages) with a postal reminder 1 month later.

This concerned information on health status, current and past jobs, and current and past occupational exposures. In the present study, only three questions were analyzed: self-rated health status (very good/good vs. moderate/poor), last occupational category for persons who had worked more than 4 months in their life (farmers, tradespeople and shopkeepers, managers and professionals, intermediate white-collar occupations, office and sales personnel, and manual workers), and a question taken from the effort/reward imbalance at work questionnaire for persons at work in date of the study [15] (“Considering all my efforts and achievements, I receive the respect and prestige I deserve at work”). The first two questions were chosen because they are global indicators of health and occupation, respectively. The last one was chosen because it is a subjective constraint at work, which could be difficult to correct for nonresponse. Because of the small partial nonresponse to these questions (<5%), item nonresponse was not treated here.

Download English Version:

<https://daneshyari.com/en/article/10513527>

Download Persian Version:

<https://daneshyari.com/article/10513527>

[Daneshyari.com](https://daneshyari.com)