OPEN

The Evolving Usefulness of the Test-negative Design in Studying Risk Factors for COVID-19

To the Editor:

In a recent paper in this journal,[1] we described how to use the test-negative design as an efficient tool for identifying risk factors for COVID-19. In the early period of the pandemic, almost all tested persons were symptomatic, which led to the proposal of using an ancillary population-based control group—in addition to the test-negative controls—because it is likely that different respiratory diseases share common risk factors.

As the COVID-19 pandemic has progressed, the testing situation has changed: more and more persons without symptoms are being tested. This obviates the necessity of an additional population control group and may make it possible to disentangle the risk factors for becoming infected with SARS-CoV-2 from the risk factors for becoming diseased with COVID-19.

The inclusion of persons without symptoms in test-negative design studies will facilitate investigation of social factors (e.g., occupation, remote work status, socioeconomic status, risk tolerance, and personal activities) that increase risks of exposure and infection, with or without subsequent development of symptoms. It will also allow comparisons of the relative effectiveness of interventions (e.g., masks, vaccines) against the different endpoints of infection and disease, which is of high relevance to decision-making regarding public policies. In contrast, test-negative design studies that are restricted to symptomatic persons will primarily bring to light biologic/clinical factors (e.g., sex, age, immunosuppression, asthma, and pregnancy) associated with developing symptomatic disease after infection.[1,2]

For the application of the test-negative design to study risk factors in situations where testing includes symptomatic as well as non-symptomatic persons, the reason for testing is important to record and account for in analysis and inference. The analysis can be stratified into three groups: participants tested due to presence of symptoms consistent with COVID-19, participants tested due to contact with a case (e.g., as a result of contact tracing), and participants tested as a precautionary measure to manage risk while engaging in activities in which close contact is inevitable, such as travel or work.

In the stratum for whom the reason for testing is "having symptoms," the interpretations from of our previous paper apply. Symptomatic cases and controls can be enrolled from the test-positives and test-negatives and compared to each other, and more can be learned by triangulating the findings with an additional population control group.[1] Adjusting for severity has been advocated because of potential noncollapsibility[3]; however, such problems are usually trivial.[4] In the stratum of persons who are tested because of a recent contact, without having symptoms, it is not necessary to add an additional population control group as such a test-negative design in fact refers to transmission risk factors in the community. For example, a test-negative design may elucidate which family member situations (e.g., parent-to-child) lead to infections more often; alternatively, if the close contact is a co-worker or friend who tested positive, the test-negative design might become more detailed in the sense of asking test-positives versus test-negatives how close the contact was, which may lead to some refinements in general precautions (e.g., meetings in open air vs. closed rooms). If the contact is instead a warning by for example a phone application, this will be unexpected, and the reasons leading to test positivity will likely not be obvious, but this may lead to an inquiry about general risk situations (i.e., participating in certain activities) of those testing positive versus negative. Similar inferences will be possible in test-negative design analyses of persons being tested as a precautionary measure.

In summary, increasingly people are being tested for a variety of reasons; it is therefore necessary to control for reason for testing in the analysis, which may enrich the application of the test-negative design as a tool for identifying risk factors for SARS-CoV-2 infection and COVID-19 disease.

Jan P. Vandenbroucke
Department of Clinical Epidemiology
Leiden University Medical Center
The Netherlands
Departments of Medical Statistics, Non-communicable Disease Epidemiology and Infectious Disease Epidemiology
London School of Hygiene and Tropical Medicine
London, United Kingdom
Department of Clinical Epidemiology
Aarhus University
Denmark
J.P.Vandenbroucke@lumc.nl

Elizabeth B. Brickley
Neil Pearce
Departments of Medical Statistics, Non-communicable Disease Epidemiology and Infectious Disease Epidemiology
London School of Hygiene and Tropical Medicine
London, United Kingdom

Christina M.J.E. Vandenbroucke-Grauls
Department of Medical Microbiology and Infection Prevention
Amsterdam UMC
The Netherlands

The first version of this manuscript had been posted on ArXiv as: 2106.03713; we have not yet posted this revised version.

The authors report no conflicts of interest.

ISSN: 1044-3983/21/332-e7
DOI: 10.1097/EDE.0000000000001438

REFERENCES

1. Vandenbroucke JP, Brickley EB, Vandenbroucke-Grauls CMJE, Pearce N. A test-negative design with additional population controls can be used to rapidly study causes of the SARS-CoV-2 epidemic. *Epidemiology*. 2020;31:836–843.
2. Pearce N, Moirano G, Maule M, et al. Does death from Covid-19 arise from a multi-step process? *Eur J Epidemiol*. 2021;36:1–9.
3. Ciocanea-Teodorescu I, Nason M, Sjolander A, Gabriel EE. Adjustment for disease severity in the test-negative study design. *Am J Epidemiol*. 2021;190:1882–1889.
4. Whitcomb BW, Naimi AI. Defining, quantifying, and interpreting "noncollapsibility" in epidemiologic studies of measures of "effect". *Am J Epidemiol*. 2021;190:697–700.

Crosswalks to Convert US Census Bureau Industry and Occupation Codes, 1980–2018

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention. Where authors are identified as personnel of the International Agency for Research on Cancer/ World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

The data used for this project are publicly available (please see the eAppendix, p. 12, for a detailed list of data sources). The crosswalks are publicly available on a GitHub repository (https://github.com/johndbeard/Crosswalks-for-industry-and-occupation-codes).

The authors report no funding and conflicts of interest.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: John D. Beard, Department of Public Health, College of Life Sciences, Brigham Young University, 4103 Life Sciences Building, Provo, UT 84602. E-mail: john_beard@byu.edu.

ISSN: 1044-3983/22/332-e8
DOI: 10.1097/EDE.0000000000001440

To the Editor:

Occupation is reflective of workers' socioeconomic status (SES) and occupational exposures and experiences.[1] Therefore, occupation has been used as a measure of SES, much like education and income, or to derive occupation-based indices of SES in many epidemiologic studies.[1] Moreover, industry and occupation can be used to generate hypotheses regarding occupational exposures and experiences associated with particular health outcomes, identify groups of workers with high burdens of particular health outcomes, and target programs, interventions, and policies to workers in industries and occupations with high disease burdens to reduce occupational illness and injury.[1] Industry and occupation information can be used to link epidemiologic studies and datasets to other datasets, such as job exposure matrices, to assign quantitative or semiquantitative estimates of occupational exposures.[2] Industry and occupation (and occupation-based SES) can be considered a(n) exposure,[3] potential confounder,[4] effect measure modifier,[5] or mediator[6] depending on the research question of interest.

Industry and occupation information is often ascertained via free text fields for epidemiologic studies or datasets used for epidemiologic research (e.g., birth and death certificates),[3] which necessitates the assignment of industry and occupation codes to the data to enable epidemiologic analyses. The US Census Bureau has developed industry and occupation codes and made them publicly available for decades.[7] Furthermore, web-based tools developed and made publicly available by the National Institute for Occupational Safety and Health can be used to assign US Census Bureau industry and occupation codes to free text fields.[8]

The US Census Bureau frequently updates their codes for industry (every 5 years) and occupation (every 8–10 years), which are used for the decennial census and surveys including the American Community Survey, Current Population Survey, and Survey of Income and Program Participation.[7] Updates are necessary because industries and occupations emerge, change, and become more or less common over time.[7] Therefore, it is important to provide researchers and professionals with crosswalks that can be used to convert older (e.g., 1980) industry and occupation codes to newer (e.g., 2018) codes. A study of usual (i.e., longest held) occupation reported on death certificates in relation to mortality from amyotrophic lateral sclerosis and Parkinson's disease provides a recent example of the utility of such crosswalks.[3]

The US Census Bureau makes publicly available their updated industry and occupation codes and instructions for using them (eAppendix, p. 6; http://links.lww.com/EDE/B874). However, they do not compile this information into files or formats that can be conveniently used for epidemiologic analyses. Therefore, we have used US Census Bureau information and followed US Census Bureau instructions to create crosswalks to convert industry codes from 1980 to 1990, 1990 to 2000, 2000 to 2002, 2002 to 2007, 2007 to 2012, and 2012 to 2017 codes. We have also created crosswalks to convert occupation codes from 1980 to 1990, 1990 to 2000, 2000 to 2002, 2002 to 2010, and 2010 to 2018 codes. Our crosswalks include industry and occupation codes for the general population of workers, military personnel, and nonpaid workers. Our crosswalks can be used to convert older industry and occupation codes to newer codes (e.g., from 1980 to 1990), but they cannot be used to convert newer industry and occupation codes to older codes (e.g., from 1990 to 1980). We are making available our crosswalks, an explanation of conversion factors, instructions for using the crosswalks, our rationale for providing two versions of each crosswalk, and notes about and descriptions of our sources (eAppendix; http://links.lww.com/EDE/B874). We hope our crosswalks will enable updated analyses of industry and occupation data within the broad spectrum of epidemiologic studies in which industry and occupation data are used.

