Journal of Clinical Epidemiology 85 (2017) 17–20

Journal of Clinical Epidemiology

When is a null finding in register-based epidemiology convincing?

Lau Caspar Thygesen*

National Institute of Public Health, University of Southern Denmark, Øster Farimagsgade 5A, DK-1353 Copenhagen

Accepted 20 February 2017; Published online 22 March 2017

Publication bias has been widely discussed [1], and both in biomedical and social sciences, it has been shown that significant or stronger results have higher likelihood of publication compared to null findings [2]. The risk of not publishing null findings may even be higher in observational studies, where data are available for multiple analyses of the same data set.

For the furtherance of science, therefore, it is important to report and publish null findings. Using register-based data or routinely collected data, several epidemiological null finding studies have been published. These studies include vaccination safety studies [3], studies of detrimental effects of induced abortion [4] and blood donation [5], exposures among children [6] or during pregnancy [7], and pharmacoepidemiological studies of detrimental effects of specific drugs [8].

These null findings are often interpreted as strong evidence against an association between proposed detrimental exposures and outcomes [4]. One reason is that they are based on nationwide registers. In the Nordic countries, unique possibilities exist for linkage of several registers to include the whole population in one study [9]. These studies are therefore less prone to selection bias because all eligible persons are available for analyses and the information on exposure is collected before information on outcome, ensuring prospective data collection [10]. But whole-population studies may be prone to information bias because exposure or outcome information may be misclassified in registers or routinely collected data. In some of the null finding studies, some biases are discussed including misclassification of exposure [3–5] and outcome [3,7] and lack of power when the exposure or outcome is rare [6,8]. Furthermore, unmeasured and residual confounding could influence register-based studies because only limited and unspecific confounder information is available [10]. Methods have been suggested to evaluate the influence of unmeasured confounding [11] but will not be further discussed in this commentary because unmeasured confounding will in most cases bias the association reported away from the null, for example, confounding by indication where medications appear to cause outcomes they are meant to prevent or healthy user bias where the healthiest continue treatment [12] and will rarely explain null findings in register-based studies.

Even though register-based data are important for evaluating associations, it is critically important to evaluate biases in such studies. Furthermore, even though the whole population is included, the possible lack of generalizability of the study results to other populations should be acknowledged.

The aim of this commentary is to discuss the potential limitations of large nationwide register-based studies reporting null findings with a focus on three questions: Even though the whole population is included, is the study large enough? Is the null finding just a result of misclassification? Is the result from one country generalizable to other countries? Based on these questions, I will present recommendations for register-based studies reporting null findings.

1. Is the study large enough?

Even though register-based studies including all persons in a whole country are large, for rare exposures and diseases, the number of exposed persons and specific outcomes may still be too low. In studies on rare exposures and outcomes, even a register-based study may be susceptible to type II error. In general, sample size considerations are informative when evaluating whether the study will be suitable for the specific research question.

The procedure of calculating sample size before conducting a study is an integral element of randomized controlled trials. In observational studies, the situation is not similar even though the STROBE statement (STrengthening the Reporting of OBservational studies in Epidemiology) recommends sample size calculation when planning a new study [13]. In studies with data already available for other purposes (such as register-based studies), it is more important to evaluate whether the results

Conflict of interest: None.

* Corresponding author. Tel.: +45-65507777.
E-mail address: lct@niph.dk

http://dx.doi.org/10.1016/j.jclinepi.2017.02.011

produced will have sufficient statistical precision to contribute substantially to the literature [13]. This could be done during the protocol development of the study, for example, as part of the application process for data access. Information on expected number of exposed persons and number of events is often available, but information on relevant effect sizes and effects of adjustment of confounding variables on the risk estimates is often difficult to include in a formal sample size calculation. The recommendation of including information on relevant effect sizes is also presented in the RECORD statement (REporting of Studies Conducted Using Observational Routinely Collected Health Data) [14].

In some studies, the lack of power or small power of the specific investigation has been discussed [6,8]. These considerations are better placed as part of the planning of studies with expected low number of exposed persons or a low number of diseased persons. This leaves the reader with better opportunity to evaluate whether the study is large enough to substantially contribute to the literature.

2. Is the null finding just a result of misclassification?

Register-based studies use reporting from clinicians, health care professionals, and administrative personnel. This leaves room for misclassification of exposure, outcome, and confounders. In general, when exposure information is collected before the event of a case, exposure will be nondifferentially misclassified, that is, misclassified independently of outcome. In the same way, outcome information may be misclassified independently of exposure because it is separated in time. According to epidemiological methodology, nondifferential misclassification will generally underestimate the association between exposure and outcome at least for relative risk measures which are the most commonly used effect measures. This bias has been discussed in some register-based null findings [3–5,7], and in general, it has been argued that this misclassification will result in an underestimation of the association. In studies reporting null findings, this means that the result could be a result of misclassification.

To illustrate the influence of nondifferential misclassification, a calculation of observed risk estimates in the presence of suboptimal sensitivity and specificity of exposure and outcome was performed. A study of 200,000 persons with 1,500 outcomes among 100,000 exposed (incidence proportion = 1,500/100,000 = 1.5%) and 500 among unexposed (0.5%) resulting in a relative risk of 3.00 was constructed. The CRAN package *episensr* was used to estimate the observed relative risk under different values of sensitivity and specificity. Because the misclassification is nondifferential, the sensitivity and specificity of outcome are similar for exposed and nonexposed and vice versa.

Table 1 presents the influence of nondifferential misclassification of exposure, showing that a sensitivity and specificity above 0.90 result only in minor bias. For some exposures, like vaccination status, which is tax refunded and payment to the general practitioner is only made when the vaccination is registered, the sensitivity and specificity are probably high. For other exposures like drug use, the information is based on prescription redemption and not the actual consumption of the drug. This could lead to decreased specificity because some patients redeeming the drug will not actually consume the drug (secondary noncompliance). In this situation, the observed relative risk will be an underestimate of the real effect.

The influence of nondifferential misclassification of the outcome of the same relative risk is heavily influenced by the specificity of the outcome, while sensitivity has only minor influence (Table 2). Even a very low sensitivity with very high specificity does not influence the association, while specificity below 0.99 introduces severe bias.

The sensitivity and specificity depend heavily on the outcome of interest. In a review of validity of specific diagnoses in the Danish National Patient Register, Schmidt et al. [16] showed that the sensitivity varied strongly between diagnoses and procedures and was very low for some diseases. In general, specificity was higher (above 0.95), but even a specificity of 0.95 could result in marked attenuation of the relative risk (Table 2).

A final note is that even though misclassification in register-based studies will generally be nondifferential

Table 1. Influence of nondifferential misclassification of exposure on observed relative risk in a cohort study of 200,000 persons with 1,500 outcomes among exposed persons (incidence proportion = 1,500/100,000 = 1.5%) and 500 among unexposed persons (0.5%)

	Specificity					
Sensitivity	**0.60**	**0.70**	**0.80**	**0.90**	**0.95**	**1.00**
0.60	1.22 (82)	1.35 (73)	1.50 (63)	1.68 (53)	1.79 (47)	1.91 (41)
0.70	1.36 (72)	1.50 (63)	1.65 (54)	1.83 (45)	1.94 (40)	2.05 (35)
0.80	1.56 (60)	1.70 (52)	1.86 (44)	2.04 (35)	2.14 (31)	2.25 (26)
0.90	1.85 (44)	2.00 (37)	2.16 (30)	2.33 (23)	2.43 (19)	2.54 (15)
0.95	2.09 (33)	2.22 (27)	2.37 (21)	2.54 (15)	2.64 (12)	2.74 (8)
1.00	2.43 (19)	2.54 (15)	2.67 (11)	2.82 (6)	2.90 (3)	3.00 (0)

Correct relative risk of 3.00. Bias % in parentheses.
Calculations performed by the CRAN package *episensr* [15].

Table 2. Influence of nondifferential misclassification of outcome on observed relative risk in a cohort study of 200,000 persons with 1,500 outcomes among exposed persons (incidence proportion = 1,500/100,000 = 1.5%) and 500 among unexposed persons (0.5%)

	Specificity					
Sensitivity	**0.90**	**0.95**	**0.98**	**0.99**	**0.999**	**1.00**
0.60	1.05 (96)	1.10 (91)	1.25 (79)	1.46 (66)	2.50 (17)	3.00 (0)
0.70	1.06 (95)	1.12 (90)	1.29 (77)	1.51 (62)	2.56 (15)	3.00 (0)
0.80	1.07 (94)	1.14 (88)	1.33 (74)	1.57 (59)	2.60 (13)	3.00 (0)
0.90	1.08 (93)	1.16 (87)	1.36 (72)	1.62 (56)	2.64 (12)	3.00 (0)
0.95	1.08 (93)	1.17 (86)	1.38 (71)	1.64 (55)	2.65 (11)	3.00 (0)
1.00	1.09 (92)	1.17 (85)	1.39 (70)	1.66 (54)	2.67 (11)	3.00 (0)

Correct relative risk of 3.00. Bias % in parentheses.
Calculations performed by the CRAN package *episensr* [15].

[10], in some situations, the outcome may be differentially misclassified, for example, if exposed persons are followed more closely by health care professionals than nonexposed persons. This could result in ascertainment bias which will overestimate the association [17].

3. Is the result from one country generalizable to other countries?

Even if researchers from one country show that an exposure is not associated with an outcome in that specific population, it is important to evaluate whether these results are valid in other populations. Other populations could be the same country at some other time period or other geographical populations [10]. A register-based study in one given time period including the whole population could be viewed as a sample of a larger theoretical population independent of time and place [18].

Generalization in epidemiology has been defined as an elaboration of scientific theory, and theories can be viewed as general statements of nature that tell us what to expect in settings where people were not studied [19]. This view of generalization is most applicable to associations that do not change over time, for example, the influence of smoking on lung cancer studied among British doctors [20] is probably generalizable to other populations at different times and in different places. Other associations, for example, socioeconomic differences in smoking prevalence, may not be generalized to other times or places.

4. Recommendations

Based on this presentation, I propose three recommendations to make null findings more convincing.

The first recommendation is to consider whether the register-based study of a whole population will provide results with sufficient precision to substantially contribute to the literature. This is especially important in studies of rare exposures or events where a study of one population may be too small to convincingly support a null finding. The effect size should be based on effect sizes that are clinically relevant or relevant to public health. The sample size consideration can be informal in a situation where data are already available through registers [13] but should be performed before the register-based study is initialized, for example, as part of the protocol development, or at least before the analyses are performed.

The second recommendation is to perform sensitivity analyses of nondifferential misclassification of exposure and outcome, which may have strongly attenuated the observed association. Such sensitivity analysis should not be a qualitative discussion of the pattern and direction of influence of misclassification but should include quantitative analyses to evaluate the magnitude of the misclassification [15]. This will add to the discussion of whether the null finding could be a result of misclassification. Different methods of quantitative sensitivity analyses of misclassification have been proposed [15], for example, by including information on misclassification of exposure, outcome, or confounders and then reconstruct the data set as it would have been in case of no misclassification. This evaluation of misclassification should be based on estimates of sensitivity and specificity from validation studies. If these studies are not available, exposures and outcomes that may relate to the variables of interest could be used, but, preferably, validation studies should be conducted.

The third recommendation is that the discussion in the papers should include a discussion of the generalization of the results. This discussion should be based on descriptives of the exposure prevalence and pattern over time, the outcome definition and incidence over time, and finally, the confounder distribution. This information is important to consider for other researchers when they evaluate whether the results from this single country are generalizable to other times and places.

Register-based whole-population studies reporting null findings have often been put forward as strong evidence against an association between proposed detrimental exposure and outcome. In this commentary, I recommend that null finding studies should include sample size considerations even though the study is based on whole population, should quantitatively evaluate whether nondifferential misclassification could explain the null finding, and evaluate the generalizability of the results to other populations.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2017.02.011.

References

[1] Ioannidis JP. Why most published research findings are false. PLos Med 2005;2(8):e124.

[2] Franco A, Malhotra N, Simonovits G. Social science. Publication bias in the social sciences: unlocking the file drawer. Science 2014; 345:1502−5.

[3] Madsen KM, Hviid A, Vestergaard M, Schendel D, Wohlfahrt J, Thorsen P, et al. A population-based study of measles, mumps, and rubella vaccination and autism. N Engl J Med 2002;347:1477−82.

[4] Melbye M, Wohlfahrt J, Olsen JH, Frisch M, Westergaard T, Helweg-Larsen K, et al. Induced abortion and the risk of breast cancer. N Engl J Med 1997;336:81−5.

[5] Edgren G, Reilly M, Hjalgrim H, Tran TN, Rostgaard K, Adami J, et al. Donation frequency, iron loss, and risk of cancer among blood donors. J Natl Cancer Inst 2008;100:572−9.

[6] Pedersen C, Raaschou-Nielsen O, Rod NH, Frei P, Poulsen AH, Johansen C, et al. Distance from residence to power line and risk of childhood leukemia: a population-based case-control study in Denmark. Cancer Causes Control 2014;25(2):171−7.

[7] Lassen J, Bager P, Wohlfahrt J, Bottiger B, Melbye M. Parvovirus B19 infection in pregnancy and subsequent morbidity and mortality in offspring. Int J Epidemiol 2013;42:1070−6.

[8] Svanstrom H, Pasternak B, Hviid A. Use of azithromycin and death from cardiovascular causes. N Engl J Med 2013;368:1704−12.

[9] Thygesen LC, Daasnes C, Thaulow I, Bronnum-Hansen H. Introduction to Danish (nationwide) registers on health and social issues: structure, access, legislation, and archiving. Scand J Public Health 2011;39(7 Suppl):12−6.

[10] Thygesen LC, Ersboll AK. When the entire population is the sample: strengths and limitations in register-based epidemiology. Eur J Epidemiol 2014;29(8):551−8.

[11] Uddin MJ, Groenwold RH, Ali MS, de Boer A, Roes KC, Chowdhury MA, et al. Methods to control for unmeasured confounding in pharmacoepidemiology: an overview. Int J Clin Pharm 2016; 38(3):714−23.

[12] Sturmer T, Jonsson Funk M, Poole C, Brookhart MA. Nonexperimental comparative effectiveness research using linked healthcare databases. Epidemiology 2011;22:298−301.

[13] Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. Int J Surg 2014;12(12):1500−24.

[14] Nicholls SG, Quach P, von Elm E, Guttmann A, Moher D, Petersen I, et al. The REporting of studies conducted using observational routinely-collected health data (RECORD) statement: methods for arriving at consensus and developing reporting guidelines. PLoS One 2015;10:e0125620.

[15] Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int J Epidemiol 2005;34:1370−6.

[16] Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol 2015;7: 449−90.

[17] Storandt M, Morris JC. Ascertainment bias in the clinical diagnosis of Alzheimer disease. Arch Neurol 2010;67(11): 1364−9.

[18] Hartley HO, Sielken RL Jr. A ‘‘super-population viewpoint’’ for finite population sampling. Biometrics 1975;31:411−22.

[19] Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol 2013;42:1012−4.

[20] Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. Br Med J 1950;2(4682):739−48.