

Does adjustment for non-differential misclassification of dichotomous exposure induce positive bias if there is no true association?

Igor Burstyn^{*,1}

Environmental and Occupational Health, Drexel University, Philadelphia, PA, United States of America

ABSTRACT

This article is a response to an off-the-record discussion that I had at an international meeting of epidemiologists more than decade ago. It centered on a concern, perhaps widely spread, that adjustment for exposure misclassification can induce a false positive result. I trace the possible history of this supposition and test it in a simulated case-control study under the assumption of non-differential misclassification of binary exposure, in which a Bayesian adjustment is applied. Probabilistic bias analysis is also briefly considered. The main conclusion is that adjustment for the presumed non-differential exposure misclassification of dichotomous does not “induce” positive associations, especially if the focus of the interpretation of the result is taken away from the point estimate. The misconception about positive bias induced by adjustment for exposure misclassification, if more clearly explained during the training of epidemiologists, may promote appropriate (and wider) use of the adjustment techniques. The simple message that can be derived from this paper is: “Exposure misclassification as a tractable problem that deserves much more attention than just a typical qualitative throw-away discussion”.

Introduction

There is a suspicion among some epidemiologists that adjustment for error in exposure estimate can artificially create an apparent positive association, a false positive. The typical argument to support this notion proceeds along these lines:

In most cases an argument is made that the observed exposures were not influenced by the knowledge of health outcome and therefore, exposure misclassification (measurement error) is typically non-differential. If non-differential exposure misclassification (measurement error) attenuates an estimate of risk, then adjusting for this phenomenon will increase the risk estimate proportionately to the presumed extent of imprecision in the observed exposure. By simply assuming an ever-increasing magnitude of error in exposure, one can arrive at a correspondingly increasing risk estimate even if there is no true association.

While there is no empirical evidence (to my knowledge) of such abuse of the adjustment methods, the obvious opportunity for a dishonest or naïve individual to bias the results seems to have cast a shadow of suspicion on all misclassification and measurement error adjustment techniques. I illustrate the reason for this suspicion of

adjustment methods being unfair in the context of the regression calibration method for adjustment of the odds ratio (OR) (classical additive measurement error) introduced by Rosner et al. [1] in Fig. 1. We can see that as the non-differential measurement error increases (or is assumed to increase) the larger is the push towards higher values in the adjusted OR away from the null. The further is the observed OR from the null, the stronger is this adjustment, but when the observed OR = 1, there is no adjustment (in practice, we almost never observe OR exactly equal to one). By simple algebra, a seemingly unimpressive observed point estimate of OR of 1.3 would be pushed to an alarming value of 3.7² if one can show (or misrepresent) that slope relating true and observed exposure is 0.2, given that all other assumptions of Rosner et al. [1] hold.

Arguments illustrated above are rarely openly voiced but may not be an uncommon concern in private discussions and are one reason put forward against adjustment for known exposure misclassification. However, if “measurement error is threatening our profession”, [2] then the apparent avoidance of exposure misclassification adjustment techniques [3,4] seems to be an unhealthy tendency. Reflecting on Willett, [5] this state of the affairs has largely not changed since the late 1980's, even though most of then-enumerated practical barriers are gone (namely: incorporation of uncertainty, accessible adjustment for commonly used regression models, lack of software), albeit there is still

* Corresponding author.

E-mail address: igor.burstyn@drexel.edu.

¹ Present address: Department of Environmental and Occupational Health, Dornsife School of Public Health, Drexel University, Nesbitt Hall, Room 614, 3215 Market Street, Philadelphia, PA 19104, USA.

² $\exp(\log(1.3)/0.2)$

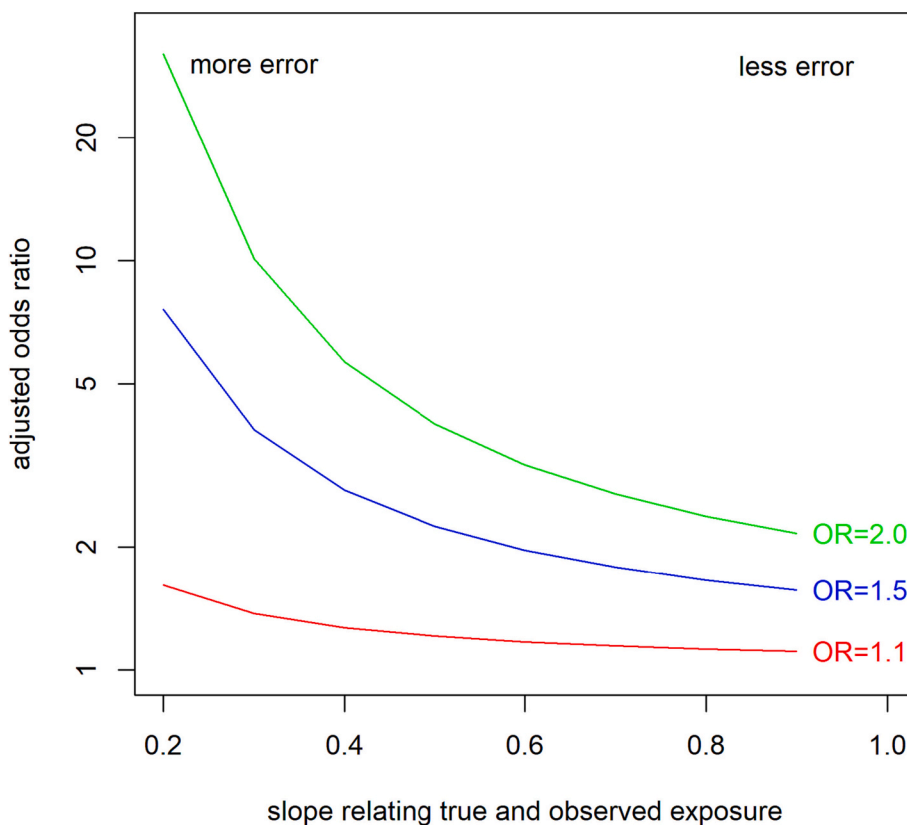


Fig. 1. The effect of adjusting odds ratio (OR) for the ever-increasing measurement error; the observed OR are given in the corresponding colors next to each line; if the wrong measurement error parameter is used, the adjusted value is biased; assuming more error leads to large adjusted OR, unless the observed OR is exactly 1; (see Rosner et al. [1] for the details of regression calibration, where the “slope” is denoted by λ).

a shortage of reliability and validity studies. The problem with lack of validation studies is self-aggravating: since there is largely no intent to adjust for misclassification and measurement error in epidemiology, reliability and validity studies are not performed (despite strong encouragement from journals like *Epidemiology*, which has a separate section devoted to validation studies³). This implies that the barriers to adjustment for measurement error in exposure hypothesized by Willett [5] were not effective targets of an intervention. Some of the most recent efforts to motivate epidemiologists to adjust for measurement error (e. g., by Wallace in 2020 [6]) appear to be using the same motivation as Willett [5] – the concerns about under-estimating effects or false negatives – and therefore are likewise likely will not stimulate any change to the epidemiologic practice. (One exception to the aforementioned gloomy assessment appears to me to be nutritional epidemiology, within which an impressive arsenal of measurement error adjustment techniques, including validation studies, has been implemented starting at least in the early 1990’s, as exemplified by [1,7–11].)

Maybe it is time to listen to the concerns of practicing epidemiologists about methods for misclassification and measurement error in

exposure advocated by statisticians? One of them may be the rarely articulated worry that adjustment for non-differential exposure misclassification can produce false positive by exaggerating true effect. After all, it has been written almost 30 years ago that non-differential misclassification of exposure does not always leads to an underestimate of risk, [12] implying that the adjustment for it should sometimes allow for the conclusion that the true effect may be weaker than the observed. Wacholder et al. [13] clarified in comment on Sorahan et al. [12] that there is a difference between misclassification process and misclassification in the resulting data: whereas the process of misclassification may be perfectly non-differential it would be difficult to expect that the empirical sensitivity and specificities will be exactly the same. In fact, there are both theoretical and practical reasons for non-differential misclassification [14] and measurement error [15] of exposure to cause false positives, likely contributing the “replication crisis”, which is habitually linked with misuse of null hypothesis testing. [15].

This article is my attempt to contribute to overcoming the reluctance of some (many) epidemiologists to explicitly tackle the problem of misclassification of dichotomous exposure. It is a follow-up to my work on multiplicative measurement error that tackled the same question in a context of a cohort study with either binary or continuous outcome, which is an unpublished (in peer-reviewed literature) prequel to the current manuscript. [16] Those readers who wish to jump to an in-depth yet accessible treatment of exposure measurement error and misclassification issues in epidemiology are advised to proceed to the publications of the STRATOS initiative (<https://stratos-initiative.org/>). [17,18]

The above perception of artificial inflation of exposure-response associations due to misclassification and measurement error adjustment can perhaps be traced to some of the early adjustment methods introduced to epidemiologists, such as the relationship popularized by

³ *Epidemiology*: Instruction for authors: “Validation Studies (2000 words) Validation studies should follow the outline for an Original Research Article and should provide estimates to inform bias analyses or otherwise be of use in epidemiologic research (see editorial). Examples include estimates of measurement error for continuous variables, classification parameters for discrete variables (sensitivity, specificity, or positive and negative predictive values), strengths of association to inform analyses of an unmeasured confounder, or participation proportions within combinations of exposures and outcomes. The validation study should be designed and the results presented to optimize their utility in other similar settings.”; <https://edmgr.ovid.com/epid/accounts/ifaauth.htm> (accessed 6/12/2023).

Table 1
Priors on the sensitivity ($SN \sim Beta(a_p, b_p)$) and specificity ($SN \sim Beta(a_q, b_q)$), their means and variances (V).

Scenario	Youden's J	Sensitivity				Specificity			
		Mean	a_p	b_p	V(SN)	Mean	a_q	b_q	V(SP)
1	0.05	0.50	5.0	5.0	0.023	0.55	5.5	4.5	0.023
2	0.15	0.55	5.5	4.5	0.023	0.60	6.0	4.0	0.022
3	0.25	0.60	6.0	4.0	0.022	0.65	6.5	3.5	0.021
4	0.35	0.65	6.5	3.5	0.021	0.70	7.0	3.0	0.019
5	0.45	0.70	7.0	3.0	0.019	0.75	7.5	2.5	0.017
6	0.55	0.75	7.5	2.5	0.017	0.80	8.0	2.0	0.015
7	0.65	0.80	8.0	2.0	0.015	0.85	8.5	1.5	0.012
8	0.75	0.85	8.5	1.5	0.012	0.90	9.0	1.0	0.008
9	0.85	0.90	9.0	1.0	0.008	0.95	9.5	0.5	0.004

Armstrong, [19] between the true and observed slopes of linear regression via the coefficient of reliability, which assumes a non-differential classical additive measurement error model. A similar relationship is known for the true and the observed relative risks [20] and ORs. [1] It is obvious from equations in [1,19,20] that as the measurement error grows larger, then these adjustment methods will yield an ever increasing *point* estimates of the association parameters. Consequently, a person wishing to “game” the rules can always postulate a non-differential measurement error sufficiently large to arrive at some target “elevated” *point* estimate of the association parameter, e.g., OR > 2 deemed by Doll as not weak. [21]

The simple approaches to the measurement error and exposure misclassification problems reviewed above do not reflect the state of the art in the field. [17,18,22,23] For example, well-established Bayesian methods that reconcile our knowledge about measurement error and misclassification with the available data do not simply scale the naïve point estimates by a multiplier derived from a validation study. [22,23] Please note that even if the adjustment is as simple as using a multiplier, false positives cannot be manufactured, because the multiplier would also be applied to endpoints of the naïve interval estimate, the null would be inside the corrected interval if and only if it is inside the naïve interval. In fact, it was shown in Bayesian framework “that failing to adjust for misclassification can (lead one to) overstate the evidence”, but “an honest admission of uncertainty about the misclassification” can result in a more accurate estimate of the association parameter; “neither of these phenomena are predicted by common rules-of-thumb”. [24] Nonetheless, epidemiological practice is dominated by the use of rules-of-thumb in dealing with measurement error and misclassification. [3,4]

My specific purpose in this article is to illustrate that odds ratios estimated from an unmatched case-control study are not materially biased upward by employing Bayesian adjustment for suspected non-differential misclassification of a dichotomous exposure. Thus, my aim is akin to that of a demonstration project that has the pedagogic value of combating a misconception about a specific way of analyzing data, not

Table 2
Summaries of posteriors odds ratios under different priors on the misclassification parameters: case-control study of 200 cases and 400 controls with apparent prevalence of exposure 20% and apparent odds ratio = 1; SN = sensitivity, SP = specificity.

Scenario	Assumed misclassification			Posterior odds ratio		
	Youden's J	mean(SN)	mean(SP)	Median	Interquartile range	
1	0.05	0.50	0.55	1.05	0.38	3.74
2	0.15	0.55	0.60	1.12	0.47	3.12
3	0.25	0.60	0.65	1.12	0.53	2.42
4	0.35	0.65	0.70	1.04	0.57	1.72
5	0.45	0.70	0.75	1.01	0.62	1.62
6	0.55	0.75	0.80	1.02	0.73	1.44
7	0.65	0.80	0.85	1.02	0.77	1.33
8	0.75	0.85	0.90	1.01	0.80	1.26
9	0.85	0.90	0.95	1.01	0.85	1.19

an attempt to trick a statistical method to perform in an unexpected manner.

Methods

I consider a case-control study with exposures that is either truly binary or a dichotomy of continuous exposures. The study, if properly analyzed, should reveal no association between the exposure and the outcomes, i.e., true OR is equal to one. Such a state of nature would yield, *in probability*, OR = 1 (equal prevalence of exposure in cases and controls). Next, I imagined that exposures among 200 cases are compared to 400 controls, assuming no important confounding or effect modification. The apparent prevalence of exposure is 20% in cases and controls, yielding a conventionally estimated frequentist OR 1.00 with 95% confidence interval 0.68 to 1.48. I would interpret such an estimate as likely excluding strong association with 95% certainty (using intuition of Doll [21]), provided that all sources of bias are so small as to be ignorable.

In keeping with typical experience in occupational epidemiology (e.g., ever exposed according to experts to diesel exhaust [25]) and assessment of stigmatized behavior (e.g., maternal smoking in pregnancy [26]), I assumed that sensitivity of exposure classification (SN) is smaller than its specificity (SP). Adopting the common assumption in the field, I dealt only with exposure misclassification that is non-differential with respect to the perfectly ascertained case status, even though I believe that this is a rare exception (an accessible argument to support this belief is repeated in Singer et al. [27], with more technical treatment as differential-due-to-dichotomization misclassification problem in the textbook book of Gustafson [23] and an even more detailed exposition in [28]).

Next, I made the common constraint of better than random exposure classification, on average: SN + SP > 1. I explore the combinations of SN and SP that are centered on 0.5 to 0.9 and 0.55 to 0.95, respectively. The uncertainty about SN and SP is captured by Beta distributions, with its parameters chosen in a manner that keeps variance of these distributions about the same across the scenarios, but more importantly, small relative to the means (i.e., these are strong Bayesian priors). The actual parameters of Beta distributions that I used as priors on SN and SP are given in Table 1.

I adjusted for misclassification using a method described in the textbook of Gustafson [23] (in section 5.6.2) and in [29] that employs Gibbs sampler. The method accounts for uncertainty about SN and SP in case-control studies and is shown to be less biased than an unrealistic (yet easy to implement) approach of using fixed values of SN and SP. The method can be extended to account for covariates. [23] I do not present technical details, but the implementation is given in Supplementary material 1 and all of my other calculations are in Supplementary material 2. Heuristically, various combinations of SN and SP are sampled from prior and used to recalculate the “true” number of exposed and unexposed cases, leading to the candidate misclassification-adjusted OR. The resulting candidate adjusted OR is more likely to be retained as a sample from the posterior distribution if it fits the data and does not rely

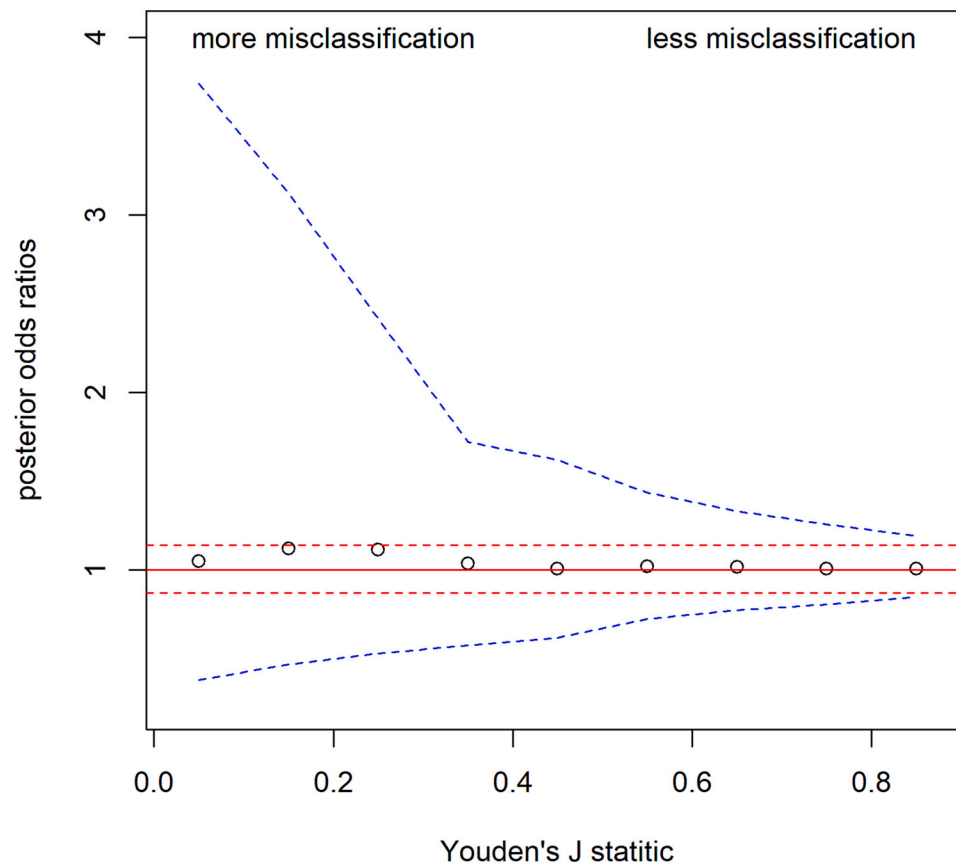


Fig. 2. Median posterior odds ratios (circles) and their interquartile range (blue dashed line) in relation to assumptions about exposure misclassifications (Youden's J statistic); the red line is the frequentist point estimate and dashed red lines denote its 75% confidence limits. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

on an extreme combination of SN, SP, and prevalence of exposure; any combinations of these parameters that lead to negative cell counts in contingency tables are excluded. The process is repeated many times till samples from posterior appear to be independent and these are then used to describe percentiles of the posterior distribution of the OR to construct credible intervals. I conducted analysis under uninformative prior on OR, which some may find unrealistic, but in this manner, I examined the effect of the manipulation of only the misclassification parameters.

All calculations were conducted in R software (R version 4.3.0 (2023-04-21 ucrt) – “Already Tomorrow” Copyright (C) 2023 The R Foundation for Statistical Computing Platform: x86_64-w64-mingw32/x64 (64-bit)). [30]

Results

I show in Table 2 that after adjustment for exposure misclassification most of the medians of the posterior of OR were shifted towards values >1 , with more extreme values associated with assumption of greater misclassification (smaller Youden's J-statistic = $SN + SP - 1$) [31]. The misclassification-adjustment algorithm found it harder to identify acceptable candidate samples from the posterior with the greater presumed misclassification (small J-statistic), signaling that it is difficult to justify adjusted OR that is far from the null when data alone does not favor such effects and is assumed to be of poor quality. I take this as one of the hedges against the risk of the priors on misclassification parameters dominating the data. This pattern is illustrated in Fig. 2. The credible intervals (the red dashed lines) are wider than confidence intervals (blue dashed lines) when exposure misclassification is high and the opposite when low. This shows how Bayesian procedure

compensates for overconfidence of its frequentist counterpart's assumption of perfect exposure assessment. Only by cherry-picking in the extremes of the posterior distribution under presumed considerable misclassification (J-statistic ≤ 0.25 here) can one conclude that these adjusted estimates supported a “strong” association, as defined by Doll [21], when data did not.

Discussion

The method for adjustment for exposure misclassification that I illustrated in this paper does not show any tendency to inflate evidence of an association that is not supported by data alone. However, the same care should be taken in interpreting such adjusted association parameters as ought to go into considering the validity of inferences drawn from naïve estimates, with the added comfort of knowing that the impact of exposure misclassification on the results has been reduced.

My simulation focused on one hypothetical study and only one method of adjustment, thus a concern about generalizability may linger. The reader is invited to use my code in the appendices to evaluate their own situation. My simulation is idealized, because the observed OR is almost never exactly equal to one and in such situations, there may be more tendency for the adjustment algorithm to produce extreme values (especially with fixed SN and SP, details not shown). But if there is no true association between exposure and the outcome, the totality of observed effect estimates, absent other biases, would center on the null, and my argument that one should not blame measurement error and misclassification adjustment procedures for false positives would hold. Other methods for handling measurement error and misclassification that are not specifically designed to adjust for the bias due to errors, e.g., quantitative (probabilistic) bias analyses popularized by Lash et al., [32]

may not have the same desirable properties as those illustrated here. [33] (I note that in the particular case examined here, quantitative bias analyses yield the simulated OR that are barely distinguishable from the naïve estimates, not even allowing for the possibility of strong effects with poor quality of exposure and thus are overly conservative; details in Supplementary material 3.)

As in all statistics, my arguments hold *in probability*: see [12,34–37] on this and other the misconceptions about measurement error and misclassification of exposure. Therefore, one can never escape the subjective nature of interpretation of any given data, because we do not have the luxury of infinite sample size and multiple replications in epidemiology. At least with the misclassification adjustment procedures, there is a theoretical guarantee that the estimates do not tend to be biased if misclassification is accurately evaluated. Hence the need for informative exposure validation studies.

In the policy realm, one can view adjusted estimates in Table 2 as proving a range of effect estimates consistent with data and models. This can be useful for determining how much risk is tolerable, given the information contained in data and model. Thus, we are not reduced to dichotomous interpretation of any associations as either present or absent. For example, if there is indeed a reason to suspect that the study is of poor quality (scenario 1), then it is perhaps wise to consider a policy that allows for the possibility of OR as high as 3.7 (75th percentile of the posterior distribution). This may stimulate epidemiologic research that is known to be more like scenario 9, i.e., of higher quality, and 75th percentile of OR at 1.2, for the same data. [38] The simple and unoriginal [39] message that can be derived from this is: do not focus on point estimates, but mind the gap between boundaries that reflect variability in the estimate.

It is also worth reiterating the cautionary note of Armstrong: “If corrections are carried out on the basis of *incorrect* information on error magnitude, bias may be increased, rather than decreased.” [19] The emphasis in the above quote in italics in the quote is mine as it reinforces the notion that only incorrect information about error in exposure will induce bias. In Bayesian methodology for measurement error and exposure misclassification, one is allowed to be uncertain about the misclassification parameters and the exact knowledge of the true distribution of exposure is not necessary. [23] I observed that the prior on the degree of misclassification did in fact influence the central tendency of the posterior distributions of the association parameters. Therefore, while blatantly incorrect assumptions about misclassification structure and magnitude are likely to lead to biased inferences, Bayesian methods appear to be able to reflect uncertainty about the magnitude of misclassification in the estimates that they yield, while still providing informative results. In other words, in the Bayesian framework an investigator no longer has to rely on correct adjustment for getting one number right, which is indeed a risky proposition.

Conclusions

Please treat exposure misclassification as a tractable problem that deserves much more attention than just a qualitative (throw-away) discussion: quantitative adjustments for exposure misclassification does not induce bias in the hands of an honest investigator.

CRedit authorship contribution statement

Igor Burstyn: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

I co-chair a tripartite (industry-government-academia) committee that is engaged in exploring and promoting greater use of epidemiology in regulatory risk assessment (The Health and Environmental Sciences Institute (HESI) Environmental Epidemiology for Risk Assessment

Committee; <https://hesiglobal.org/environmental-epidemiology-for-risk-k-assessment/>). I admit to having been engaged in advocacy for the use of adjustment for imperfections of exposure variables in research. I declare no other conflicts of interest.

Acknowledgements

Dr. Aline Talhouk prepared the computer code (for another project) that is included in Supplementary material 1. I am thankful to Dr. Neal Goldstein for his insightful critique of the draft manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gloepi.2023.100132>.

References

- [1] Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *StatMed*. 1989;8(9):1051–69.
- [2] Michels KB. A renaissance for measurement error. *IntJEpidemiol*. 2001;30(3):421–2.
- [3] Jurek AM, Maldonado G, Greenland S, Church TR. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. *EurJEpidemiol*. 2006;21(12):871–6.
- [4] Burstyn I. Occupational epidemiologist's quest to tame measurement error in exposure. *Global Epidemiol* 2020;2.
- [5] Willett W. An overview of issues related to the correction of non-differential exposure measurement error in epidemiologic studies. *Stat Med* 1989;8(9):1031–40. discussion 71–3.
- [6] Wallace M. Analysis in an imperfect world. *Significance*. 2020;17(1):14–9.
- [7] Kaaks R, Riboli E, Esteve J, van Kappel AL, van Staveren W. Estimating the accuracy of dietary questionnaire assessments: validation in terms of structural equation models. *Stat Med* 1994;13(2):127–42.
- [8] Kaaks R, Plummer M, Riboli E, Esteve J, van Staveren W. Adjustment for bias due to errors in exposure assessments in multicenter cohort studies on diet and cancer: a calibration approach. *Am J Clin Nutr* 1994;59(suppl):245S–50S.
- [9] Kaaks R, Ferrarri P. Dietary intake assessments in epidemiology: can we know what we are measuring? *Ann Epidemiol* 2006;16(5):377–80.
- [10] Ferrari P, Kaaks R, Riboli E. Variance and confidence limits in validation studies based on comparison between three different types of measurements. *J Epidemiol Biostat* 2000;5(5):303–13.
- [11] Daures JP, Gerber M, Scali J, Astre C, Bonifacj C, Kaaks R. Validation of a food-frequency questionnaire using multiple-day records and biochemical markers: application of the triads method. *J Epidemiol Biostat* 2000;5(2):109–15.
- [12] Sorahan T, Gilthorpe MS. Non-differential misclassification of exposure always leads to an underestimate of risk: an incorrect conclusion. *Occup Environ Med* 1994;51(12):839–40.
- [13] Wacholder S, Hartge P, Lubin JH, Dosemeci M. Non-differential misclassification and bias towards the null: a clarification. *Occup Environ Med* 1995;52(8):557–8.
- [14] Burstyn I, Yang Y, Schnatter AR. Effects of non-differential exposure misclassification on false conclusions in hypothesis-generating studies. *Int J Environ Res Public Health* 2014;11(10):10951–66.
- [15] Loken E, Gelman A. Measurement error and the replication crisis. *Science* 2017;355(6325):584–5.
- [16] Burstyn I. Does adjustment for measurement error induce positive bias if there is no true association? arXiv 2009. 0902.1193v1 [stat.AP].
- [17] Shaw PA, Gustafson P, Carroll RJ, Deffner V, Dodd KW, Keogh RH, et al. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 2—more complex methods of adjustment and advanced topics. *Stat Med* 2020;39(16):2232–63.
- [18] Keogh RH, Shaw PA, Gustafson P, Carroll RJ, Deffner V, Dodd KW, et al. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1—basic theory and simple methods of adjustment. *Stat Med* 2020;39(16):2197–231.
- [19] Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup Environ Med* 1998;55(10):651–6.
- [20] Armstrong BG. The effects of measurement errors on relative risk regression. *Am J Epidemiol* 1990;132:1176–84.
- [21] Doll R, Sirr. Weak associations in epidemiology: importance, detection, and interpretation. *J Epidemiol* 1996;6(4):S11–20.
- [22] Gilks WR, Richardson S, Spiegelhalter D. Markov chain Monte Carlo in practice 1996. Chapman & Hall/CRC Press; 1996.
- [23] Gustafson P. Measurement error and misclassification in statistics and epidemiology 2004. Chapman & Hall/CRC Press; 2004.
- [24] Greenland S, Gustafson P. Accounting for independent nondifferential misclassification does not increase certainty that an observed association is in the correct direction. *AmJEpidemiol*. 2006;164(1):63–8.

- [25] Burstyn I, Gustafson P, Pintos J, Lavoue J, Siemiatycki J. Correction of odds ratios in case-control studies for exposure misclassification with partial knowledge of the degree of agreement among experts who assessed exposures. *Occup Environ Med* 2018;75(2):155–9.
- [26] Burstyn I, Kapur N, Shalapay C, Bamforth F, Wild TC, Liu J, et al. Evaluation of the accuracy of self-reported smoking in pregnancy when biomarker level in an active smoker is uncertain. *Nicotine Tob Res* 2009;11(6):670–8.
- [27] Singer AB, Daniele Fallin M, Burstyn I. Bayesian correction for exposure misclassification and evolution of evidence in two studies of the association between maternal occupational exposure to Asthmagens and risk of autism Spectrum disorder. *Curr Environ Health Rep* 2018;5(3):338–50.
- [28] Gustafson P, Le Nhu D. Comparing the effects of continuous and discrete covariate mismeasurement, with emphasis on the dichotomization of mismeasured predictors. *Biometrics*. 2002;58(4):878–87.
- [29] Gustafson P, Le ND, Saskin R. Case-control analysis with partial knowledge of exposure misclassification probabilities. *Biometrics*. 2001;57(2):598–609.
- [30] Team RDC. R: A language and environment for statistical computing. ISBN 3-900051-07-0.2006. Vienna, Austria: R Foundation for Statistical Computing; 2006.
- [31] Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3(1):32–5.
- [32] Lash TL, Fink AK. Semi-automated sensitivity analysis to assess systematic errors in observational epidemiologic data. *Epidemiology*. 2003;14.
- [33] MacLehose RF, Gustafson P. Is probabilistic bias analysis approximately Bayesian? *Epidemiology*. 2012;23(1):151–8.
- [34] Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. *IntJEpidemiol*. 2005;34(3):680–7.
- [35] Correction to "misconceptions about the direction of Bias from nondifferential misclassification". *Am J Epidemiol* 2022;191(12):2123.
- [36] Yland JJ, Wesselink AK, Lash TL, Fox MP. Misconceptions about the direction of Bias from nondifferential misclassification. *Am J Epidemiol* 2022;191(8):1485–95.
- [37] van Smeden M, Lash TL, Groenwold RHH. Reflection on modern methods: five myths about measurement error in epidemiological research. *Int J Epidemiol* 2020; 49(1):338–47.
- [38] Phillips CV. The economics of 'more research is needed'. *Int J Epidemiol* 2001;30 (4):771–6.
- [39] Phillips CV, LaPole LM. Quantifying errors without random sampling. *BMC Med Res Methodol* 2003;3.