Global Epidemiology 7 (2024) 100144

Contents lists available at ScienceDirect

Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology

Commentary

Does recall bias explain the association of mood disorders with workplace harassment?

Igor Burstyn [a,*], Jean-Michel Galarneau [b,c], Nicola Cherry [b]

[a] Department of Environmental and Occupational Health, Drexel University, Philadelphia, PA, USA
[b] Division of Preventive Medicine, University of Alberta, Edmonton, Alberta, Canada
[c] Faculty of Kinesiology, University of Calgary, Calgary, Alberta, Canada

ARTICLE INFO

Keywords
Information bias
Differential misclassification of exposure
Anxiety
Depression
Harassment
Bullying
Quantitative bias analysis
Probabilistic bias analysis
Occupational

ABSTRACT

Purpose: To determine the contribution of recall bias to the observed excess in mental ill-health in those reporting harassment at work.

Methods: A prospective cohort of 1885 workers in welding and electrical trades was contacted every six months for up to 5 years, asking whether they were currently anxious or depressed and whether this was made worse by work. Only at the end of the study did we ask about any workplace harassment they had experienced at work. We elicited sensitivity and specificity of self-reported bullying from published reliability studies and formulated priors that reflect the possibility of over-reporting of workplace harassment (exposure) by those whose anxiety or depression was reported to be made worse by work (cases). We applied the resulting misclassification models to probabilistic bias analysis (PBA) of relative risks.

Results: We observe that PBA implies that it is unlikely that biased misclassification due to the study subjects' states of mind could have caused the entire observed association. Indeed, the results demonstrated that doubling of risk of anxiety or depression following workplace harassment is plausible, with the unadjusted relative risk attenuated with understated uncertainty.

Conclusions: It seems unlikely that risk of anxiety or depression following workplace harassment can be explained by the form of recall bias that we proposed.

Introduction

Many studies demonstrate poor mental health in those reporting workplace bullying [1,2], but attribution of causality is not straightforward. An ideal design would be longitudinal, with measures of mental health before and after the bullying events and independent verification that the event could "reasonably be expected to cause offence, humiliation or other physical or psychological injury or illness to an employee" [3]. The difficulties of conducting such a study are substantial. Objective assessment of psychiatric conditions is time consuming, while obtaining verification from a third party is conceptually as well as logistically difficult [4]. Harassment may be evident only to the perpetrator and the victim; a report from the bullied worker may be the only credible source of information. However, studies using the worker for information about both bullying and mental health may be subject to biased reporting. A worker who is anxious or depressed, perhaps for reasons unrelated to work, may be more likely to recall or interpret workplace frictions as harassment than a worker in a happier state of mind. It is essential to keep in mind that there is no universally agreed upon definition of workplace harassment. The experience is subjective and thus the measurement must either be subjective reporting or inferences from proxy variables, with the limitations inherent in either of those. In purely technical terms, the concern is that the specificity of recall of workplace harassment may depend on respondent's frame of mind at the time of both the event itself and its reporting. When recall is affected by mental state at reporting, the appropriate conceptualization seems to us to be that of specificity being better for those who enjoy good mental health, as they report events more "objectively". However, when reporting is dependent on the state of mind at the time of the event, then those in a better frame of mind may be less likely to perceive events as harassing (indeed, some events that might be harassing for someone in a different frame of mind are arguably not

* Corresponding author at: Department of Environmental and Occupational Health, Dornsife School of Public Health, Drexel University, Nesbitt Hall, Room 614, 3215 Market Street, Philadelphia, PA 19104, USA.
E-mail address: igor.burstyn@drexel.edu (I. Burstyn).

https://doi.org/10.1016/j.gloepi.2024.100144
Received 8 February 2024; Received in revised form 25 April 2024; Accepted 26 April 2024
Available online 28 April 2024

harassment for someone in the right mood) and tend to have a better mental health at the time of reporting the outcome. This would produce confounding; the extent to which this mechanism is at play would clearly depend on the severity of the event, with the higher severity less likely to be colored by the participants' perception. The issue of whether the events that would be characterized by an "omniscient observer" as harassment regardless of the events' perception by the participants is important to consider, as that would lead to conceptualization of both sources of bias described above as a measurement error problem. Setting these important difficulties aside, for simplicity of exposition and generalization to other topics, we hereafter treat the problem, from the mathematical perspective, as that of imperfect specificity that is differential with respect to the outcome. However, the matter of confounding by mental health at the time of reporting of the outcome is addressed in passing, because it was the main feature of our previous analyses of this matter [5].

The mechanism by which a person's mental health may affect their recall and reporting of harassment, may also relate to their willingness to report anything at all, manifesting in a refusal to participate in such a study if they are enjoying positive frame of mind and do not believe that they were harassed at work. Thus, if a person retained in the cohort were more likely to suffer from mental illnesses and have experienced workplace harassment, then such a selection mechanism would lead to a spurious association between the two [6]. The bias would occur even if recall was perfect, i.e. no exposure misclassification. Mitigating against this possibility is the high participation rate in this cohort and the fact that it was not assembled with a focus on mental illness [5]. Consequently, we do not treat selection/collider bias in this manuscript as an important source of uncertainty.

Our aim was to determine the extent to which it was plausible to discount as recall bias previously observed excesses in mental ill-health in those reporting harassment[5]. Galarneau et al. [5] conducted a "simple" deterministic calculation to show that recall bias was an unlikely explanation of the association, and we now seek to examine this conclusion via a more comprehensive probabilistic (stochastic) analysis.

Methods and materials

Study design

A cohort of 1885 women and men in welding and electrical trades was assembled and contacted every six-months for up to 5 years [7]. At each contact the worker was asked, among other health and employment questions, whether they were currently anxious or depressed and whether this was made worse by work [5]. They were not asked about harassment during these periodic contacts but, at the end of the study, completed a final questionnaire to elicit "self-labelling" information on harassment through detailed questions about their experience of workplace psychological or sexual harassment or physical violence while in their trade. Respondents were asked: "during your apprenticeship were you 1) ever subjected to psychological harassment; 2) ever subjected to physical violence; 3) ever subjected to sexual harassment?", with parallel questions for the post-apprenticeship period. Any positive responses were classified as evidence of exposure to workplace bullying or harassment in the current analysis, and full details examining various aspects of the association were reported in Galarneau et al. [5].

The questionnaire that covered a wide range of topics, with a section on workplace harassment (see supplemental materials of [5]), was completed by 75% of the cohort and among those 97% also completed the Hospital Anxiety and Depression scale [8], asked later in the same questionnaire. Those who were excluded due to missing data on either the exposures or the outcomes were on average 2 years younger, male (59 vs. 52%), and in welding rather than electrical trades (62 vs. 50%). This led to an analytical sample of 1187 who had completed the harassment questions and completed at least one periodic mental health report while in their trade[5]: of these 480 (40%) reported anxiety or depression made worse by work (referred to as "cases" below). Among 531 reporting harassment (referred to as "exposed" below), 278 (52%) prospectively reported anxiety or depression made worse by work at least once. Among 656 not reporting harassment, 202 (31%) prospectively reported anxiety or depression made worse by work. The observed relative risk (RR) for anxiety or depression made worse by work in those harassed was 1.7, 95% confidence interval (CI): 1.5, 2.0.

To better appreciate the potential impact of recall bias, we provide a collection of deterministic calculations of the theoretical impact of adjustment of the observed RR of 1.7; for simplicity, we will perform calculations with odds ratios (OR) assuming prevalence of exposure of 5.00% in referents and 8.25% in cases, leading to the corresponding OR = 1.7; as this greatly simplifies the calculations without loss of generality in this specific case. If we assume that sensitivity was perfect, but the specificity was differential for the reasons we identify of 0.96 in cases and 0.98 in referents, then the true OR would 1.5, whereas with non-differential misclassification, the true OR would have been 2.2 with specificity of 0.98, and 4.4 with specificity of 0.96. An additional illustration is that if we assume that there is a null effect of the exposure, but the differential misclassification is as in the previous sentence (but the OR was calculated assuming perfect measurement, as it usually is) then an estimate of the causal effect of 1.4 would be reported. The reader is cautioned against picking fixed inputs as we have done for this illustration and implicitly declaring them to be exactly right by reporting the corrected OR estimate as if it were certain. This reprises part of the usual error of assuming no measurement error and reporting as if that assumption were exactly right. However, the calculations are very sensitive to assumptions ingrained in the input values, as is known be the case for misclassification adjustment in general[9]. Consequently, there is no substitute for the sort of quantitative bias analyses that we pursue in the manuscript.

Overview of the adjustment methodology

We will begin by articulating a model of exposure misclassification that captures the notion of recall bias by making specificity differential with respect to the outcome. We will then proceed to elicit plausible distributions of misclassification parameters from published literature using a previously developed Bayesian procedure that extracts information about validity of classification schemes from measures of agreement, which is suitable for our work given that there is no gold standard for the exposure of interest in our particular case. Once we have insights into the distribution of misclassification parameters, we will plug these into readily accessible probabilistic bias analyses algorithms. This mashing together of Bayesian and non-Bayesian methods may seem inelegant and yet it achieves our goal of adjusting for recall bias (and associated random errors) using readily available, published, methods. (We will leave development of a fully Bayesian method for future work.)

Model of recall bias and misclassification

Drews and Greenland [10] conceptualized recall bias as a problem of differential misclassification of exposure, and Moradzadeh et al.[11] implemented a Bayesian adjustment for it. We posit the following model of *misclassification of self-reported harassment*: sensitivity (SN_{01}) is the same in cases and referents (but allowed to vary by case status in each simulation), but specificity in cases (SP_1) is lower than the specificity in referents (SP_0). Exposure classification is better than chance. SN and SP in cases and referents are positively correlated. Lacking any evidence to the contrary, we assumed that exposure misclassification was consistent across confounders and independent of them.

It is essential to note that once we admit uncertainty about misclassification parameters (as we must, because these are known only vaguely), it becomes impossible to anticipate the outcome of adjustment for recall bias even though our intuition tells us that we will be reducing positive bias. Remember that our work is motivated by the supposition that recall bias may have created a non-causal association. However, there is also a random component to recall bias, e.g. it may not be of the same degree for each respondent and there are also genuine uncertainties about its magnitude in any realistic adjustment. Adjustment of purely random misclassification errors in exposure would pull the effect estimate towards greater, not smaller values. The systematic and random components of recall bias will pull the adjusted estimate in different directions, and it is not obvious until you run the numbers which might have a bigger effect.

We assumed bias from confounding and error in the outcome to be ignorable. Adjustment for prior postulated potential confounders (anxiety and depression at the time of reporting harassment, sex, trade, and total reports made)[5] decreased the estimate of RR to 1.5, 95%CI: 1.3 to 1.8 (details in **Appendix A**). It is important to clarify why anxiety and depression at the time of report of exposure were treated as confounders. There is a good argument to be made that someone who is in a poorer frame of mind is more likely to self-report harassment not because they misclassify it, but because the same experience that someone else might not consider harassment is harassment for them. It is not mismeasured but is rather a form of confounding, not measurement error. The contribution of confounding to RR appears to be in the range of 10%, an order of magnitude smaller than change in effect estimate that may typically trigger a qualitative difference in conclusions[12].

Elucidation of misclassification parameters from Conway et al.[13]

Structured questionnaires on negative events at work have been developed and studies undertaken to validate "cut points" for bullying, using "self-labelling" of bullying as the gold standard [13–15]. Of the three studies, only two reported results in sufficient detail to reconstruct contingency tables of agreement between approaches[13,15] enabling us to estimate the joint distribution of sensitivity and specificity. Conway et al.[13] evaluated recall over 12 months while Hutchison et al.[15] ascertained only current bullying. We chose Conway et al. [13] for prior elucidation as the approach is most similar to our history of bullying. Conway et al. [13] compared assessment of harassment via the short form of the Negative Actions Questionnaire (S-NAQ) with self-labelled occasional bullying at work. Both measures were self-reported retrospectively and collected at the same time. As such, we perceive them to be measures of reliability of self-reported bullying, under the assumption that neither approach can be judged superior to the other.

We followed Burstyn et al.[16] in deriving a joint prior on SN and SP from [13]. The approach relies on the mathematical relationship between SN and SP (which we wish to estimate) and results of a reliability study captured by Cohen's κ and characterized by prevalence of "exposure" or presence of a binary trait that is being accessed. Heuristically, we posit a lower plausible bound of κ and plausible exposure prevalence to calculate lower bounds of SN and SP. We then randomly draw SN and SP from the range between their lower bound and theoretical maximum values of 1. Some of these values are not mathematically compatible with the posited upper bound of κ and "exposure" prevalence (from the observed reliability study) and are therefore discarded. The result is a collection of pairs of SN and SP that are consistent with the observed reliability study and with the anticipated prevalence of "exposure" classified in the reliability study. We extracted information tabulating S-NAQ of $\geq$12 or $<$ 12 against the self-labelling report of occasional workplace bullying (**Appendix B**, **Table B1**). The two agreed with a Cohen's κ of 0.216 (95% CI, 0.190 to 0.243, standard error 0.014). Based on this, we allowed for the prior on κ to follow a uniform distribution, U(l,h), with the lower and upper bounds themselves following uniform distributions $l \sim U(0.1, 0.15)$ and $h \sim (0.2, 0.25)$. Information about agreement allows us to infer a prior on SN and SP, provided that we can also elicit a prior on true prevalence of workplace harassment in the population studied by Conway et al. [13]. We believe that literature supports a prevalence between 10 and 25% [17]. We expressed this as Beta(15.316, 76.159) with mean 16.7% and 95% CI of the mean 9.9 to 25% using R's *beta.buster* function [18]. With the three distributions described above as inputs, we obtained by simulation samples of joint prior values of SN and SP and their marginal percentiles; R[19] code is in **Appendix B**.

Priors that capture both recall bias and misclassification of exposure

We assumed that differences in specificities that are characteristic of recall bias are bounded within the range of their values that occurred in a sample that includes persons with and without anxiety or depression disorders, as in Conway et al.[13]. In developing priors, we use percentiles of SN and SP derived from the application of the methodology detailed above. We assumed that prior on SP_0 follows a triangular distribution bounded by the 25th and 97.5th percentiles of SP, with the mode at the 75th percentile. For SP_1, we adopted a triangular distribution prior bounded by the 2.5th and 97.5th percentiles of SP, with the mode at the 25th percentile, ensuring that on average $SP_1 < SP_0$, which captures the postulated recall bias. We adopted a trapezoidal distribution on SN_{01} bounded by 2.5th and 97.5th percentiles of SN, with the level between the 25th and 75th percentiles. The triangular distribution is a special case of the trapezoidal distribution, so we refer to these as "trapezoidal priors". We altered the "trapezoidal" priors to have a prior on SN that mimics its marginal distribution, with peak at the mode and bounded by the minimum and maximum of the marginal distribution: we refer to this as "triangular priors". We next considered uniform distributions with the same bounds as the "trapezoidal" priors. Next, to preserve some of the dependence of SN and SP in the prior, we fixed SN_{01} near its mean and selected SP_0 and SP_1 from the extremes of SP simulated at the mean SN. We constructed priors as uniform distributions, recognizing that uncertainty in the estimates is in the second decimal place ($\pm$0.01), such that prior on $SN_{01} \sim U(\text{mean(SN)}-0.01, \text{mean(SN)} + 0.01)$, etc., see **Appendix B**). We call this prior "fixed at mean SN".

Probabilistic bias analysis (PBA)

In our specific analysis, PBA computes a collection of "if … then…" simulated RRs by re-arranging the observed two-by-two table that yielded the original (naïve) RR using a range of misclassification parameters specified above. By moving persons between cells of the two-by-two table according to the probabilities implied in the priors, new simulated point estimates of RR and their standard errors are produced. This accounts for systematic errors due to exposure misclassification. We further account for random errors by sampling RR from each of their distributions characterized by the point estimate and its standard error. The resulting RR reflects both random and systematic errors. The procedure is repeated many times, and the distribution of simulated RR's is examined to capture what the causal effect estimate would have been if the study had not been subject to specified recall bias. PBA was carried out using the *episensr* package in R (https://CRAN.R-project.org/package=episensr) via the *probsens* function, detailed in Lash et al. [20] (Appendix C). We conducted 50,000 simulations to obtain medians and 95% simulation intervals (SI) of RR that reflect both specified random and systematic errors. Simulated tables that violated constraints of positive cell counts were discarded.

Results

We obtained 8543 samples of the joint distribution of SN and SP (Fig. 1) (marginal distributions: **Appendix B**, Fig1. B). They allowed us to estimate the following marginal 2.5%, 25%, 50%, 75% and 97.5% percentiles of SN as 0.29, 0.56, 0.75, 0.89, and 0.99, and of SP as 0.50, 0.67, 0.77, 0.88, and 0.98, respectively. The mean SN was 0.71 and at that value SP spanned the range from 0.68 to 0.87 (used to define "prior fixed at mean SN").

Consequently, results for our main analysis (trapezoid priors) are:

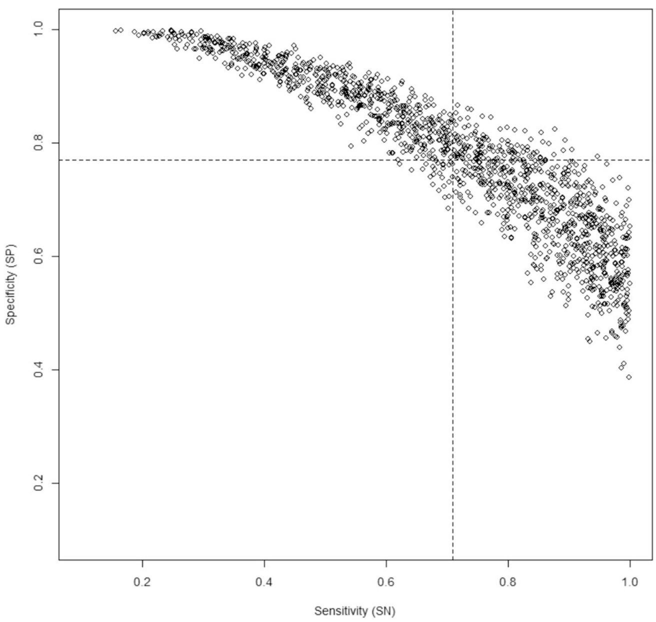

Fig. 1. Joint prior on sensitivity and specificity derived from the reliability study of self-reported workplace harassment by Conway et al. [13]; dashed lines denote means; the marginal distributions are shown in Appendix B, Fig. 1B.

$SN_{01} \sim$ Trapezoidal(a = 0.29, b = 0.56, c = 0.89, d = 0.99), $SP_0 \sim$ Triangle(a = 0.67, b = 0.98, c = 0.88), and $SP_1 \sim$ Triangle(a = 0.50, b = 0.98, c = 0.68). When considering triangular priors, we adopted $SN_{01} \sim$ Triangle(a = 0.2, b = 0.99, c = 0.95) instead, reflecting a skewed distribution with the mode close to the maximum, a near-triangular shape of the estimated marginal density function (**Appendix B**, Fig. 1B). When relaxing assumptions about misclassification parameters, we obtain $SN_{01} \sim U(0.29, 0.99)$, $SP_0 \sim U(0.67, 0.98)$, and $SP_1 \sim U(0.50, 0.98)$. We considered two levels of correlation of SN and SP by outcome: 0.4 and 0.8 (commonly recommended defaults[20]): it is reasonable to expect that misclassification parameters in cases are not independent of values in referents.

The results of the four PBA are presented in Table 1, demonstrating that, on average, a doubling of risk is plausible, with the original (unadjusted) RR of about 1.7 attenuated, on average, relative to the RR observed after accounting for exposure misclassification. However, the uncertainty in the magnitude of the effect was understated in the unadjusted analysis. The PBA seemed to be robust to choices of our "trapezoidal" vs. "triangular" priors, with the "triangular" priors leading to more precise yet weaker, on average, effects; both approaches yielded a similar proportion of simulations that had to be discarded due to violations of assumptions. The "uniform" priors had a higher rate of rejection of simulations than either "trapezoidal" or "triangular", signaling that uniform priors are less consistent with the observed data and assumed models. The priors "fixed at mean SN" yielded estimates that were on average greater than the observed RR (2.0 vs. 1.7). None of the simulations with the priors "fixed at mean SN" were discarded, which is an artefact of unrealistically assuming next to no uncertainty in the misclassification parameters. The PBA that we a priori selected as being more realistic (trapezoidal prior with the stronger correlation) yielded evidence in favor of the positive association (2.5th percentile >1.00), if one were to apply conventional considerations of the 95% uncertainty interval excluding the null.

Table 1
Probabilistic bias analysis (PBA) to account for recall bias and exposure misclassification relating risk of anxiety or depression made worse by work to self-labeling of workplace harassment in a cohort of 1187 workers in welding and electrical trades [21]; 50,000 simulations before discarding samples that violate assumptions of the analysis; observed relative risk (RR) for anxiety or depression made worse by work in those harassed was 1.7, 95% confidence interval (CI): 1.5, 2.0 and it is this estimate that is further adjusted in PBA.

Prior distributions	Correlation	Samples (%) discarded	Simulated relative risk (percentile)		
			50	2.5	97.5
Trapezoidal	0.4	31	1.9	0.6	19
	0.8	30	2.0	1.0	16
Triangular	0.4	27	1.7	0.6	12
	0.8	25	1.8	1.0	11
Uniform	0.4	45	1.9	0.4	17
	0.8	42	2.0	0.8	15
Fixed at mean SN	0.4	0	2.0	1.7	2.3
	0.8	0	2.0	1.7	2.3

Discussion and conclusions

The effect of recall bias on estimates of the effects of workplace harassment on mental health is a challenging area of research. Future validation studies might leverage methods developed in nutritional epidemiology employing multiple error-prone methods [22,23]. The correlation of each instrument with the latent construct of true exposure can be inferred under some simplifying assumptions [24]. Defining a gold standard would be a barrier, with alloyed gold standard the only attainable goal. This would be sufficient for calibrations of effect estimates to account for errors in exposure [25].

In interpreting our findings, it is important to recall that a PBA does not adjust for misclassification but produces a collection of plausible alternatives that reflect random errors while adjusting for systematic ones, without distinguishing between those that fit the data well or poorly [26]. PBAs can also yield overly imprecise estimates, specifically because it does not discard some of the less likely "adjustments". To do better, one could employ Bayesian methods as with the form of prior we employed in another setting [27].

Our method has general applicability to quantify the impact of recall bias and exposure misclassification simultaneously. While assuming perfect exposure classification leads to a precise estimate of about a doubling of risk, further accounting for plausible recall bias indicates that data are more consistent with a doubling of risk that may lie, with 95% certainty, between an implausible reduction in risk and a ten-fold increase. As foreshadowed in the introduction, there are two adjustments competing with each other in that adjustment for imperfect exposure misclassification pulls the estimate away from the null towards the true value when sensitivity and specificity are similar for cases and non-cases (the distributions of misclassification parameters by case status overlap) and downward away from the inflated naïve estimate when simulated misclassification parameters are "strongly" non-differential with respect to the outcome in the manner that we specified. Surprisingly, the net effect of adjustment for recall bias is to pull the estimate away from the null because of uncertainty about differences in sensitivity and specificity by case status. More could be learned from validation studies of measurement of workplace harassment, e.g. by stratifying such studies on some measures of mental health at the time of exposure assessment.

We believe that the current work is sufficient to demonstrate that a positive association between harassment reported at a later date and earlier periodic reports of anxiety and depression made worse by work [5] is unlikely to be validly attributed to recall bias.

CRediT authorship contribution statement

Igor Burstyn: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jean-Michel Galarneau:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation. **Nicola Cherry:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gloepi.2024.100144.

References

[1] Verkuil B, Atasayi S, Molendijk ML. Workplace bullying and mental health: a Meta-analysis on cross-sectional and longitudinal data. PloS One 2015;10(8):e0135225.
[2] Rudkjoebing LA, Bungum AB, Flachs EM, Eller NH, Borritz M, Aust B, et al. Work-related exposure to violence or threats and risk of mental disorders and symptoms: a systematic review and meta-analysis. Scand J Work Environ Health 2020;46(4):339–49.
[3] Canada Go. Requirements for Employers to Prevent Harassment and Violence in Federally Regulated Workplaces. Available from: https://www.canada.ca/en/employment-social-development/programs/workplace-health-safety/harassment-violence-prevention.html#; 2020.
[4] Cowie HNP, Rivers I, Smith PK, Pereira B. Measuring workplace bullying. Aggress Violent Behav 2002;7(1):33–51.
[5] Galarneau JM, Durand-Moreau Q, Cherry N. Reported harassment and mental ill-health in a Canadian prospective cohort of women and men in welding and electrical trades. Ann Work Expo Health 2024;68(3):231–42.
[6] Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology 2003;14(3):300–6.
[7] Cherry N, Arrandale V, Beach J, Galarneau JF, Mannette A, Rodgers L. Health and work in women and men in the welding and electrical trades: how do they differ? Ann Work Expo Health 2018;62(4):393–403.
[8] Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the hospital anxiety and depression scale. An updated literature review. J Psychosom Res 2002;52(2):69–77.
[9] Gustafson P, Le ND, Saskin R. Case-control analysis with partial knowledge of exposure misclassification probabilities. Biometrics 2001;57(2):598–609.
[10] Drews CD, Greeland S. The impact of differential recall on the results of case-control studies. Int J Epidemiol 1990;19(4):1107–12.
[11] Moradzadeh R, Mansournia MA, Ghiasvand R, Baghfalaki T, Nadrian H, Holakouie-Naieni K. Impact of age at menarche on breast Cancer: the assessment of recall Bias. Arch Iran Med 2019;22(2):65–70.
[12] Maldonado G, Greenland S. Simulation study of confounder-selection strategies. AmJ Epidemiol 1993;138(11):923–36.
[13] Conway PM, Hogh A, Nabe-Nielsen K, Grynderup MB, Mikkelsen EG, Persson R, et al. Optimal cut-off points for the short-negative act questionnaire and their association with depressive symptoms and diagnosis of depression. Ann Work Expo Health 2018;62(3):281–94.
[14] Notelaers GE, S.. The world turns at 33 and 45: Defining simple cutoff scores for the Negative Acts Questionnaire–Revised in a representative sample. Eur J Work Organ Psy 2013;22(6):670–82.
[15] Hutchinson M, Bradbury J, Browne G, Hurley J. Determining the optimal cut-off scores for the workplace bullying inventory. Nurse Res 2017;25(3):46–50.
[16] Burstyn I, de Vocht F, Gustafson P. What do measures of agreement (kappa) tell us about quality of exposure assessment? Theoretical analysis and numerical simulation. BMJ Open 2013;3(12):e003952.
[17] Nielsen MB, Matthiesen SB, Einarsen S. The impact of methodological moderators on prevalence rates of workplace bullying. A meta-analysis. J Occupation Organizat Psychol 2010;83(4):955–79.
[18] Stevenson MNT, Heuer C, Marshall J, Sanchez J, Thornton R, Reiczigel J, et al. epiR: An R package for the analysis of epidemiological data. R package version 0.9–69. 2015.
[19] Team RDC. R: A language and environment for statistical computing. ISBN 3-900051-07-0. Vienna, Austria: R Foundation for Statistical Computing; 2006.
[20] Lash TL, Fox MP, Fink AK. Applying quantitative Bias analysis to epidemiologic. Data: Springer; 2021.
[21] Galarneau J-M, Durand-Moreau Q, Cherry N. Reported harassment and mental-ill health in a Canadian prospective cohort of women and men in welding and electrical trades. medRxiv 2023;2023. 11.07.23298222.
[22] Kaaks R, Ferrari P. Dietary intake assessments in epidemiology: can we know what we are measuring? Ann Epidemiol 2006;16(5):377–80.
[23] 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.
[24] Davis ME. Structural equation models in occupational health: an application to exposure modelling. OccupEnvironMed 2012;69(3):184–90.
[25] Spiegelman D, Schneeweiss S, McDermott A. Measurement error correction for logistic regression models with an "alloyed gold standard". AmJ Epidemiol 1997;145(2):184–96.
[26] MacLehose RF, Gustafson P. Is probabilistic bias analysis approximately Bayesian? Epidemiology 2012;23(1):151–8.
[27] 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.