Measurement error in medical research: a systematic review of current practice

Timo B. Brakenhoff, MSc[1], Marian Mitroiu, MSc[1], Ruth H Keogh, PhD[2], Karel G.M. Moons, PhD[1], Rolf H.H. Groenwold, MD, PhD[1], Maarten van Smeden, PhD[1]

1. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, the Netherlands

2. Department of Medical Statistics, London School of Hygiene and Tropical Medicine, U.K.

Corresponding author:

T. B. Brakenhoff, MSc. (ORCID: 0000-0003-3543-6296)

Julius Center for Health Sciences and Primary Care

University Medical Center Utrecht

PO Box 85500, 3508 GA Utrecht, the Netherlands

T: +31 88 756 9618; E: T.B.Brakenhoff-2@umcutrecht.nl

ABSTRACT

In medical research, covariates (e.g. exposure and confounder variables) are often measured with error. While it is well accepted that this introduces bias and imprecision in exposure-outcome relations, it is unclear to what extent such issues are currently considered in research practice. The objective was to study common practices regarding covariate measurement error via a systematic review of general medicine and epidemiology literature. Original research published in 2016 in 12 high impact journals was full-text searched for phrases relating to measurement error. Reporting of measurement error and methods to investigate or correct for it were quantified and characterized. 247 (44%) of the 565 original research publications reported on the presence of measurement error. 83% of these 247 did so with respect to the exposure and/or confounder variables. Only 18 publications (7% of 247) used methods to investigate or correct for measurement error. Consequently, it is difficult for readers to judge the robustness of presented results to the existence of measurement error in the majority of publications in high impact journals. Our systematic review highlights the need for increased awareness about the possible impact of covariate measurement error. Additionally, guidance on the use of measurement error correction methods is necessary.

Key Words: bias; epidemiology; measurement error; medicine; misclassification; review

WHAT'S NEW

- About half of the reviewed original research from 12 top-ranked general medicine and epidemiology journals mentioned the concept of measurement error in some form.
- Investigations into the impact of covariate (exposure and confounder) measurement error on studied relations as well as the application of measurement error correction methods were rare.
- This extensive systematic review confirms suspicions raised over a decade ago by many authors as well as another review on a similar topic: that the potential impact of measurement error on studied relations is often ignored and misunderstood.
- Consequently, it is difficult for readers to judge the robustness of presented results to the existence of measurement error in the majority of publications in high impact journals.
- Our systematic review highlights the need for both, increased awareness about the possible impact of covariate measurement error, as well as guidance on the use of measurement error correction methods.

1. Introduction

Measurement error is one of many key challenges to making valid inferences in biomedical research [1]. Errors in measurements can arise due to inaccuracy or imprecision of measurement instruments, data coding errors, self-reporting, or single measurements of variable longitudinal processes, such as biomarkers. With the increased use of data not originally intended for research, such as routine care data, ‘claims’ databases and other sources of ‘big data’, it is conceivable that measurement error is becoming increasingly prevalent in this field [2].

It is generally well accepted that measurement error and classification error (hereinafter collectively referred to as measurement error) in either the dependent variable (hereinafter *outcome*) or independent explanatory variables (hereinafter *covariates*; e.g. exposure and confounder variables) can introduce bias and imprecision to estimates of covariate-outcome relations. Among others, several textbooks [3–6], methodological reviews [7,8] and a tool-kit [9], have demonstrated how to examine, quantify, and correct for measurement error in a variety of settings encountered in epidemiology. Most of this work has been focused on measurement error in covariates given its conceived greater impact on studied relations than measurement error in the outcome [4]. Despite these resources, it is suspected that the attention it receives in applied medical and epidemiological studies is insufficient [10,11].

Over a decade ago, a review of 57 randomly selected publications from three high ranking epidemiology journals reported that 61% of the reviewed publications recognized the potential influence of measurement error, but only 28% made a qualitative assessment of its impact on their results, and only one quantified its potential impact on results [12]. In light of the increasing prevalence of measurement error in medical and epidemiological research and increasing availability of methods and software to account for measurement error, a new and more comprehensive investigation into current practice is necessary.

We conducted a systematic review to quantify the extent to which (possible) measurement error in covariates is addressed in recent medical and epidemiologic research published in high impact journals. To guide the understanding of the results of the review, we briefly introduce key concepts in the field of measurement error.

2. Measurement error

Many variables of interest in medical research are subject to measurement error. Instead of an error-free and unobserved, *true* value of a variable, researchers have to deal with an imperfectly measured, *observed* value. For the remainder of this section, we consider the erroneous measurement and perfect measurement of a single underlying entity as different variables. Examples of variables prone to measurement error include the long-term average level of a variable biological process (such as blood pressure) when the researcher may only have access to a single measurement; average daily caloric intake measured using food frequency questionnaires; diabetic status ascertained using electronic health record data; and individual air pollution exposure based on measurements from a fixed monitor.

In the context of multivariable statistical models, such as regression models, measurement error can be present in the outcome and/or covariates. We focus on error in covariates. In their seminal text-book, Carroll et al. [5] describe the effect of measurement error in covariates as a "triple whammy": covariate-outcome relationships can be biased, power to detect clinically meaningful relationships is diminished, and features of the data can be masked. Whether bias is present, and if so its direction and magnitude, depend on the form of the measurement error. It is therefore important to quantify any bias due to measurement error and to obtain corrected estimates where possible. Three important considerations in this process are: identification of the variables of interest that are measured with error, what type of measurement error is present, and what additional information is available to help characterize the error.

2.1 Types of measurement error and their effects

Measurement error is characterized differently for continuous and categorical variables. For continuous variables, four types of error can be distinguished that describe how the observed variable relates to the unobserved, true variable.

The simplest type of measurement error, *classical* error, occurs when the observed variable can be expressed as the true variable plus a random component with zero mean and constant variance. As a result, when measurements of an observed variable (e.g. blood pressure) are repeatedly taken from the same person, the average of these measurements would approach that person's true variable value (e.g. the usual blood pressure level) as the number of replicate measurements increases. In the context of etiologic research, the estimated exposure-outcome relation will be biased towards the null (also known as attenuation) when only the exposure variable is measured with classical error [5]. However, the estimated relations between the confounders (provided that they are measured without error) and the outcome in the same model could be biased in either direction, depending on the form of the relation between the main exposure and the confounders. It follows that classical measurement error in one or multiple confounders can result in bias in either direction for the exposure-outcome relation, even if the exposure is measured without error [13]. The direction and magnitude of this bias is thus unpredictable and this holds for different regression models of interest in epidemiology, including logistic, Cox and linear regression models [5].

Two other types of error that are related to the classical error model are *systematic* and *differential* error. When the error is systematic, the observed variable is a biased representation of the true variable and the average of repeated observed measurements would no longer approach the true variable value. Measurement error is described as ‘differential’ if the mismeasured covariate would help predict the studied outcome even if the values on the true covariate would have been observed (i.e., the error is dependent on the outcome, conditional on the values of the true covariate). Differential error depending on the outcome can arise when the outcome occurs prior to the measurement of covariates, as in case-control studies. Both systematic and differential error can cause bias in the exposure-outcome, or more generic, the covariate-outcome relation in either direction.

The last common type of measurement error is called *Berkson* error, which arises when the true variable is equal to the observed variable plus a random component with zero mean and constant variance; i.e. the true and observed variable reverse roles, compared to classical error. Berkson error can occur when group averages are used in place of individual measurements. Examples of Berkson error are often found in environmental epidemiology where individual exposure to air pollutants is set equal for individuals that live within a certain radius of an air pollution monitor. While Berkson error in covariates can diminish precision, in many cases it does not cause bias in the estimates of the exposure-outcome relation [5,14].

For categorical variables, measurement error is commonly referred to as *misclassification*. Misclassification can be summarized using sensitivity and specificity when the variable is binary. In the situation where a single binary exposure is related to an outcome, random non-differential misclassification present in the exposure will result in attenuation of this exposure-outcome relation [1]. However, when the exposure has more than two categories, when the exposure is subject to systematic or differential misclassification, or when confounders measured with error are added to the analysis model, it is once more difficult to predict in which direction the estimate of the true exposure-outcome relation will be biased [4].

2.2 Measurement error correction methods

Several methods have been proposed that aim to correct for bias due to measurement error in covariates. We highlight a few measurement error correction methods below that can be used when continuous variables are measured with error. The methodological literature addressing measurement error corrections is extensive, e.g. [1,4,5,14].

Regression calibration was proposed by Rosner, Willett and Spiegelman in 1989 [15]. The essence of regression calibration is that the observed error-prone covariate is replaced by a prediction of the expected value of the true variable in the analysis. Regression calibration can be used when there is non-differential classical or systematic measurement error. This approach requires information on the degree of measurement error, which is the error variance in the case of classical error. We note how this information can be obtained below.

Cook and Stefanski proposed the simulation-extrapolation (SIMEX) method [16]. This method works via a two-step procedure. First, data are simulated by adding additional error of different magnitudes to the observed exposure measurements; the simulated data sets are used to estimate the effect of this additional error on the exposure-outcome relation. As a second step, the estimate of the exposure-outcome relation is extrapolated back to the situation where there is no measurement error using an extrapolation model which relates the estimated exposure-outcome association parameter to the degree of measurement error. Like regression calibration, this method requires information about the amount of measurement error (variance) in the observed variable. SIMEX as described above assumes non-differential classical error, yet has also been extended to deal with misclassified categorical variables [17].

Alternatively, a large range of so-called latent variable models have been suggested to account for measurement error during analysis. Latent variable models generally rely on replicate measurements of error-prone measures to estimate a latent variable to represent the true error-free variable [18]. This latent variable can replace the observed error-prone variable in the exposure-outcome analysis or can be modelled directly in the exposure-outcome model, for instance, using Structural Equation Modeling [18,19].

We acknowledge that it can be very challenging to determine the structure and amount of measurement error due to the plethora of underlying (unobserved) factors that may influence it. While further guidance is required on how to assess the amount and type of measurement error in practice, it can generally be recommended to collect additional data, whenever feasible, either in a subset of the study sample or possibly in an external validation sample, to compare observations on a covariate that is (suspected of being) measured with error and an error free representation of that covariate (if such a ‘gold standard’ exists). This information can subsequently be used to study measurement error structures, amount of measurement error, and to inform measurement error correction methods (e.g. regression calibration or SIMEX, among others), which allow for a measurement error corrected analysis on the whole study sample. Alternatively, when available, repeated measurements of a covariate measured with error can be used to quantify measurement error variance and allow for measurement error corrected analyses.

2.3 Availability of additional information for measurement error corrections

Additional information about the form of the measurement error is often required to quantify its impact on the exposure-outcome relation and potentially correct for it. This information can be obtained from validation data or, if the error is classical, replicate measurements.

Validation data contains the error-prone variable alongside the true variable. Typically, these data are only available for a subset of the study sample or the information may come from an external source, such as another data set or published results. For example, when participants of a study have been requested to self-report their BMI via an online questionnaire (the error-prone variable), a subset may have had their BMI measured according to a systematic protocol by a research assistant (the 'true' variable).

Replicate measurements may consist of multiple measurements with error from the same instrument (e.g. multiple measurements of blood pressure), or sometimes multiple measurements from different instruments that aim to measure the same true variable (e.g. multiple diagnostic tests for the same disease). Replicates may be observed for all or a subset of study participants and is often collected when measuring a variable biological process.

When validation or replication data are acquired from external sources, the similarity of these research settings with the current setting, i.e., *transportability,* needs to be assessed [5].

If there is little information available to inform measurement error correction methods or to assess the structure of the measurement error model, the potential impact of measurement error can still be explored through sensitivity analyses. Hypothetical scenarios can then be assessed by rerunning the analysis assuming fixed amounts of measurement error or misclassification. A formal extension of sensitivity analysis, referred to as "probabilistic sensitivity analysis" (thoroughly detailed by Greenland & Lash in chapter 19 of [1]) can also be used to assess many potential scenarios with differing amounts of measurement error simultaneously, and obtain an estimate of the exposure-outcome relation adjusted for both systematic and random errors.

3. Methods

We performed a systematic review of original research published in 2016 in high-impact medical and epidemiological journals. Our aims were to: i) quantify and characterize the reporting of measurement error in a main exposure and/or confounder variables and their possible impact on study results and ii) quantify and characterize the use of available methods for investigating or correcting for measurement error in the exposure and/or confounder variables.

Using the Thomson Reuters InCites rankings of 2015 [20], the 6 highest-ranking journals in the categories "General & Internal Medicine" (New England Journal of Medicine, Lancet, JAMA, BMJ, Annals of Internal Medicine and JAMA Internal Medicine) and "Epidemiology" (International Journal of Epidemiology, European Journal of Epidemiology, Epidemiology, American Journal of Epidemiology, Journal of Clinical Epidemiology, Journal of Epidemiology and Community Health) were identified. The journal Epidemiology Review was excluded as it is an annual journal. All publications of the above-mentioned journals from the period 01/01/2016 to 31/12/2016 were identified using PubMed (see search string in Appendix A).

Title and abstracts were screened by one reviewer (TB). Publications that were not original research (e.g. brief reports, essays, cohort profiles, and guidance papers) were excluded. Also excluded were: methodological research, review and meta-analysis research, qualitative research, policy oriented studies, descriptive studies, studies that analyzed data on an aggregated level, and publications that did not assess individual health related exposures and outcomes.

After initial screening, a full-text search was performed in the remaining manuscripts using a Boolean search with stemming in Adobe Acrobat XI Pro. The search string contained the term “measurement error” and synonyms such as “misclassification” or “mismeasured”, as well as phrases relating to the validity of the collected data, including “information bias” or “self-reported”. The exact search string can be found in Appendix B. Manuscripts that contained any of the terms included in the search string were screened to assess whether they: a) discussed measurement error with respect to previous studies or the design of the current study; b) discussed the potential of measurement error in one or more of the covariates; c) discussed the potential effect of measurement error on the presented study results; or d) described methodology to investigate or correct for any measurement error. Publications that fulfilled at least one of these criteria were included in the following data extraction step.

The included publications were reviewed independently by two readers (TB and MM) using a standardized data extraction form (see Appendix C). This form was pilot tested by four researchers (TB, MS, RG, MM). Disagreements were discussed until consensus was reached. The elements extracted included: design of data collection, study characteristics, clinical domain, characterization of variable(s) subject to measurement error (exposure/confounder), sections of the article where measurement error was mentioned (abstract/introduction/methods/results/discussion), reporting of possible effects of measurement error on study results (direction and magnitude of effect), reporting of the assumed type of error, reporting of methods that investigated the impact of, or attempted to correct for, measurement error in exposure or confounder variables.

Articles that reported impact of measurement error or corrections for measurement error were included for additional review by four readers (TB, MS, RG, MM). For these publications, data were extracted from the main document and the supplementary materials. The methods used were characterized, alongside how this was reported and the type of additional information used.

4. Results

Figure 1 depicts the number of included papers at each step of the review process. Of the 1178 articles found in PubMed, 565 (337 from Epidemiology journals and 228 from General & Internal Medicine journals) were judged as original research satisfying our inclusion criteria. Of these, 247 (44%) directly addressed measurement error in some form. Characteristics of these included studies are found in Table 1. Eighteen of these publications (3% of the 565) investigated the possible impact of, or corrected for, measurement error. Thirteen of these eighteen publications were from Epidemiology journals (4% of the 337 Epidemiology publications) and the remaining five were from General & Internal Medicine Journals (2% of the 228 General & Internal Medicine publications). Table 2 shows from which journals the publications that directly addressed measurement error originated.

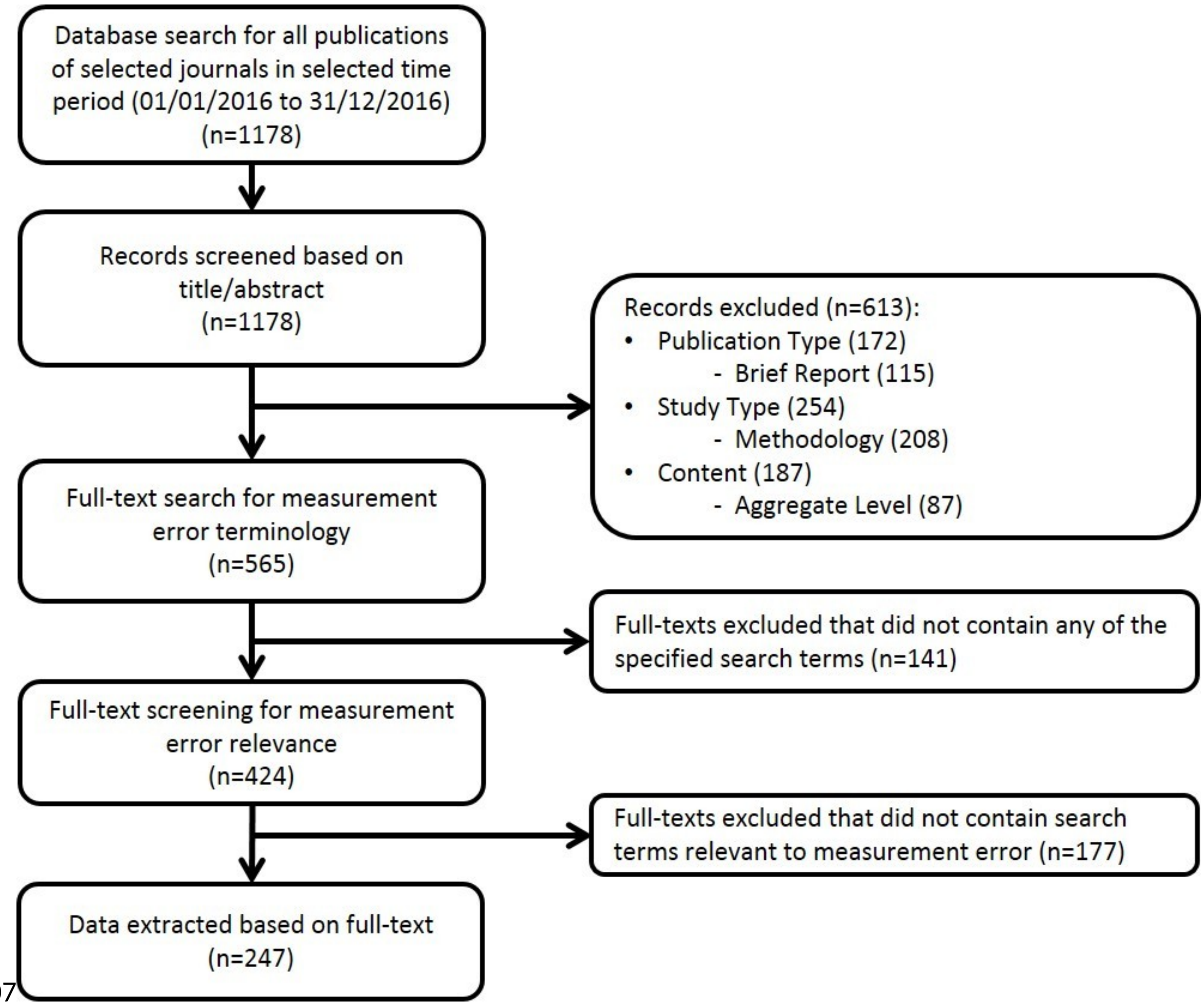

Fig. 1 Flow Diagram Detailing the Systematic Review Process

Table 1 General Characteristics of the 247 Publications That Explicitly Report on Measurement Error (ME) in Some Form.

Characteristic	No. of Studies	% of 247
ME in which variable		
Exposure	195	79
Confounder	44	18
Outcome	115	47
Exposure & Confounder	35	14
ME discussed in which section		
Abstract	8	3
Introduction	22	9
Methods	49	20
Results	9	4
Discussion[a]	219	89
ME in previous study[b]	88	36
ME prevented by design[c]	60	24

ME = Measurement error

[a] 174 (70%) publications considered ME **only** in the discussion section

[b] Mentions made of ME pertained to previously published research and not to the study presented in the published paper.

[c] ME in the presented study was prevented due to decisions made during the design of the study.

Table 2 In Which Journals the 247 Publications That Reported on Measurement Error (ME) and That Investigated or Corrected for it Were Published.

Journal Name	Publications that reported on ME		Publications that investigated/corrected for ME (n=18)
	No.	% of 247	
Am J Epidemiol	60	24	2
Ann Intern Med	7	3	1
BMJ	30	12	1
Epidemiology	17	7	4
Eur J Epidemiol	23	9	2
Int J Epidemiol	50	20	4
J Clin Epidemiol	2	1	0
J Epidemiol Community Health	37	15	1
JAMA	2	1	1
JAMA Intern Med	16	6	2
Lancet	2	1	0
N Engl J Med	1	0.5	0

ME=Measurement error

4.1 Measurement error in main exposure variables

A total of 195 (79%) of the 247 publications reported on (possible) measurement error in the main exposure variable. Of these 195, 89 (46%) reported the presence of measurement error in the exposure but did not mention, or were unclear about, its possible effect on the studied relations; 66 (34%) reported that the measurement error in the exposure did or could have led to underestimation of the exposure–outcome relation; 25 (13%) reported that measurement error in the exposure was anticipated to have had no or a negligible effect on the estimated exposure-outcome relation; three (2%) publications stated that measurement error in the exposure could have led to both over- or underestimation of the studied effect; and one publication reported a possible overestimation of the exposure–outcome relation. 11 (6%) publications explicitly reported that their exposure variable was measured *without* error.

Information about the nature of measurement error was reported by 59 (30%) of the 195 publications. For instance, these papers made general statements about the structure of the measurement error (e.g. using terms such as “random error” or “differential error”) or provided details on possible dependence of the measurement error on other variables in the analysis. Four publications (3%) were specific about the assumed error model; one publication assumed the error to be of the Berkson type and the remaining three investigated the form of the measurement error.

4.2 Measurement error in confounder variables

Of the 44 publications that reported on measurement error in the confounders, 29 (66%) reported the presence of measurement error without mentioning (or were unclear about) its possible effect on the studied relations, six (14%) reported that the measurement error in the confounder did or could have led to underestimation of the relation between the main exposure and the outcome, and four (9%) reported that measurement error in the confounder was anticipated to have no or only a negligible effect on the main exposure–outcome relation. None of the publications reported on possible overestimation of the main exposure-outcome relation due to confounders measured with error. Five (11%) publications explicitly reported that their confounder variable(s) were measured *without* error.

Six (14%) of the 44 publications made general statements about the structure of the measurement error. One discussed the assumed error model.

4.3 Measurement error impact and correction

Of the 247 publications that directly reported on measurement error, 18 (7%) either investigated its impact on the studied relations or corrected the exposure-outcome relation for measurement error (Table 3).

Table 3 Characteristics of the 18 Publications That Reported on Investigation of or Correction for Measurement Error (ME).

Characteristic	No. of Studies	% of 18
Study design		
Cohort	14	78
Case-control	4	22
Exposure field		
Lifestyle/Health (not nutrition)	9	50
Nutrition	1	6
Environment	3	17
Education	1	6
Medical intervention	4	22
ME in which variable		
Exposure	15	83
Continuous	6	
Categorical	9	
Confounder	1	6
Continuous	1	
Categorical	0	
Exposure & confounder	2	11
Both categorical	1	
Continuous & categorical	1	
How was ME dealt with		
Regression calibration	2	11
Latent variable analysis	2	11
Application specific methods*	3	17
Sensitivity analysis	11	61

ME=Measurement error
*Methods designed specifically for a field of applied research

Seven publications (39%) of the 18, applied measurement error correction methods. Two publications used regression calibration, relying on internal validation data. One of these [21] used additional data gathered for a subset of participants to account for measurement error in the exposure (daily coffee intake). The other [22] corrected for measurement error in several anthropomorphic measurements using data from earlier validation studies conducted within the same cohort. One publication [23] used a non-parametric method [24] to correct for underestimation of the exposure-outcome relation because of assumed random measurement error in the exposure (plasma triglycerides values at baseline). Another publication [25] used external observed air quality monitoring data to correct their estimates of individual air pollutant exposure. Two publications used factor analysis to define a latent exposure. One [26] implemented a latent variable model to determine each individual's disability score using many different items of a conceptual framework for describing functioning and disability. This score was then used in a regression analysis. In another [27] the factor analysis was embedded in a structural equation model where latent PTSD status was estimated from multiple clusters of symptoms suggestive of PTSD. Finally, Leslie et al. [28] used an ad-hoc approach, coined 'least significant change', to take into account inherent instrument measurement error when ascertaining exposure status (absolute bone mineral density difference).

The remaining 11 (61%) of the 18 publications investigated the impact of measurement error on the exposure-outcome relation using sensitivity analyses. In five publications [29–33], an assumption was made about the amount of possible measurement error and its effect on the exposure-outcome relation was quantified. Often this was achieved by looking at a subgroup of the original sample for which the mismeasured variable of interest was assumed to be measured with less or no error. Four publications [34–37] looked at multiple scenarios in which they assumed different amounts of measurement error. The remaining two publications [38,39] performed a probabilistic sensitivity analysis. All authors reported that the results of the sensitivity analyses were either similar to those of the conventional analyses or did not influence their conclusions. No study investigated the impact of measurement error on their results using an external dataset.

5. Discussion

This review provides an overview of the attention given to measurement error in recent epidemiological and medical literature. We found that a high proportion (44%) reported on the (possible) presence of measurement error in one or more recorded variables. 70% of these addressed measurement error in a qualitative manner only in the discussion section. In contrast, few publications (7%) used some form of measurement error analysis to investigate or correct the exposure-outcome relation for the presence of measurement error in covariates.

The results of our review can be compared to the 2006 review by Jurek et al. [12]. In their review of 57 papers published in 2001 in 3 high impact epidemiology journals (American Journal of Epidemiology, Epidemiology and the International Journal of Epidemiology), the authors reported that 61% discussed measurement error in exposure variables in some form. Based on the 565 original research publications included in our review, we found the attention given to exposure measurement error in 2016 to be lower (35%). In both studies, roughly half of included papers did not report on the expected impact of measurement error on the studied relations (2001: 51% vs 2016: 46%), and the application of measurement error correction methods was found to be relatively rare (2001: 9% vs 2016: 3%). However, a marked difference was found in the proportion of papers reporting possible attenuation of the exposure-outcome relation due to measurement error (2001: 9% vs 2016: 34%). We note that the comparison between the reviews should be interpreted with some caution due to differences in the designs of the reviews. For instance, our review was based on a larger sample of publications, examined measurement error in confounder variables, and considered both “General & Internal Medicine” and “Epidemiology” journals.

Half of the 565 included publications in our study reported about measurement error being present in any of the studied variables. In our opinion, this proportion is quite high considering the denominator includes studies in which measurement error may not be an issue (e.g. clinical trials with objective endpoints such as mortality). As such, many authors justifiably ignored the issue and did not report on it in the final publication.

As compared to the abundance of qualitative statements made about the presence of measurement error, we found formal measurement error evaluations to be surprisingly rare. About 4% of the papers that made a qualitative statement about measurement error quantified its impact using sensitivity analyses. Only 2% used formal measurement error correction methods. Several reasons for this low prevalence can be postulated. In practice it can be very challenging to properly assess the structure and amount of measurement error. Obviously, determining a strategy to account for measurement error in the analysis is then very difficult. But even when a suitable strategy can be determined and data are available to implement the strategy, there may still be lack of familiarity with these methods and available software among applied researchers, medical readers and journal editors, which may frustrate the adoption of these methods in the medical literature. For example, statistical software such as R [40] can be used to implement regression calibration (see supplementary material of [9]), SIMEX [41] and latent variable modeling [42]. There also seems to be a lack of educational materials and courses that provide guidance for practicing researchers, peer-reviewers and editors on how to use, assess and interpret results from measurement error correction methods.

A need for better understanding of measurement error in medical and epidemiologic research is further supported by a noticeably high incidence (about one third of those that discussed exposure measurement error) of manuscripts which claimed underestimation of the exposure-outcome relation due to measurement error. This conclusion was supported by a claim that the error was non-differential in about a third of the publications. Besides the fact that the non-differential measurement error assumption was regularly made without proof and is easily violated [14], non-differential measurement error also does not guarantee attenuation of the studied relation towards the null. As discussed in section 2, even classical (random) error can result in bias away from the null in several likely scenarios, e.g. when multiple variables in the analysis model are measured with error or when an exposure variable has more than two categories. In recent decades, several authors have attempted to dispel the myth that exposure measurement error always leads to attenuation of the studied relation [43–45].

Of the 18 publications that investigated or corrected for measurement error, most manuscripts reported both the original ('naïve') and the measurement error corrected results. Unfortunately, descriptions of the used methods were often not provided. Indeed, half of the publications that performed sensitivity analyses reported the results using only a single line in the results section claiming similarity of results to the main analysis (e.g., [36]). A similar proportion of these publications also only investigated one possible measurement error scenario.

Our review has some limitations. It cannot be ruled out that our full-text search strategy may have missed papers that mentioned measurement error. Although our search string covered a broad range of terminology related to measurement error, papers using a-typical terms may have been overlooked. This might have led to an underestimation of the number of publications that discussed measurement error. This limitation is unlikely to have a substantial impact on the estimated percentages and conclusions, given that the intention was to give a general impression of current practice of measurement error reporting. Second, in our review we ignored measurement error issues related to the outcome variable. While measurement error in outcome variables is often assumed to pose less problems than measurement error in covariates [4], we acknowledge that this choice limits our findings. Finally, there are measurement errors that influence analyses that do not strictly fall in the multivariable (exposure – outcome) classification. Specifically, diagnostic test accuracy studies often suffer from measurement error in the disease verification procedure, a problem known as “absence of gold standard”, and were outside the scope of this review. Reviews of methods [46,47] and the use of methods [48] to account for disease verification problems are found elsewhere.

Our systematic review also has strengths. By using modern, automated full-text searching capabilities in Adobe Reader, a comprehensive review could be conducted with about 10 times as many included publications as the earlier review conducted by Jurek et al. [12] . We were able to consider all publications from 12 top-ranked journals for a full one-year period. This full-text searching approach is likely to be much more sensitive than common search strategies that are limited to wording in the title or abstract. In addition, the full-text procedure allowed us to systematically pinpoint the article section in which references to measurement error were made.

In conclusion, we found that measurement error is often discussed in high impact medical and epidemiologic literature. However, only a small portion proceeds to investigate or correct for measurement error. Renewed efforts are required to raise awareness among applied researchers that measurement error can have a large impact on estimated exposure-outcome relations and that tools are available to quantify this impact. More guidance and tutorials seem necessary to assist the applied researchers with the assessment of the type and amount of measurement error as well as the steps that can subsequently be taken to minimize its impact on the studied relations. Given the unpredictable nature of the impact of measurement error on the studied results, we advise authors to report on the potential presence of measurement error in recorded variables but exercise restraint when speculating about the magnitude and direction of its impact unless the appropriate analysis steps are taken to substantiate such claims. Also, we recommend authors to make more use of available correction methods and probabilistic sensitivity analyses to correct analyses for variables that were measured with error. Given the increasing use of data not originally intended for medical or epidemiological research, we anticipate that the use and understanding of measurement error analyses and corrections will become increasingly important in the near future.

FUNDING

This work was supported by the Netherlands Organization for Scientific Research (NWO-Vidi project 917.16.430 granted to R.H.H. Groenwold).

CONFLICT OF INTEREST

Conflicts of interest: none

REFERENCES

[1] Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. 3rd ed. Philadelphia, PA, USA: Lippincott Williams & Wilkins; 2008.

[2] Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. N Engl J Med 2016;375:1216–9. doi:10.1002/aur.1474.Replication.

[3] Fuller WA. Measurement Error Models. John Wiley & Sons; 1987.

[4] Gustafson P. Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments. Boca Raton, United States: Chapman and Hall/CRC; 2004.

[5] Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. 2nd ed. Chapman & Hall /CRC Press; 2006.

[6] Buonaccorsi J. Measurement Error: Models, Methods and Applications. CRC Press; 2010.

[7] Stefanski LA. Measurement Error Models. J Am Stat Assoc 2000;95:1353–8.

[8] Guolo A. Robust techniques for measurement error correction: a review. Stat Methods Med Res 2008;17:555–80. doi:10.1177/0962280207081318.

[9] Keogh R, White I. A toolkit for measurement error correction, with a focus on nutritional epidemiology. Stat Med 2014;33:2137–55. doi:10.1002/sim.6095.

[10] Buzas JS, Stefanski LA, Tosteson TD. Measurement Error. In: Ahrens W, Pigeot I, editors. Handb. Epidemiol., 2014, p. 1241–82. doi:10.1007/978-0-387-09834-0.

[11] Blackwell M, Honaker J, King G. A Unified Approach to Measurement Error and Missing Data: Overview and Applications. Sociol Methods Res 2015:1–39. doi:10.1177/0049124115589052.

[12] Jurek AM, Maldonado G, Greenland S, Church TR. Exposure-measurement error is

frequently ignored when interpreting epidemiologic study results. Eur J Epidemiol 2006;21:871–6. doi:10.1007/s10654-006-9083-0.

[13] Brakenhoff TB, van Smeden M, Visseren FL, Groenwold RHH. Random measurement error: why worry? An example of cardiovascular risk factors. PLoS One 2018;In Press.

[14] Ahrens W, Pigeot I, editors. Handbook of Epidemiology. 2nd ed. New York, USA: Springer-Verlag New York; 2014.

[15] Rosner B, Willett W, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med 1989;8:1051–69.

[16] Cook J, Stefanski L. Simulation-extrapolation estimation in parametric measurement error models. J Am Stat Assoc 1994;89:1314–28. doi:10.2307/2290994.

[17] Küchenhoff H, Mwalili SM, Lesaffre E. A general method for dealing with misclassification in regression: The misclassification SIMEX. Biometrics 2006;62:85–96. doi:10.1111/j.1541-0420.2005.00396.x.

[18] Skrondal A, Rabe-Hesketh S. Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. Crc Press; 2004.

[19] Kline RB. Principles and practice of structural equation modeling. Guilford publications; 2015.

[20] Thomson Reuters. InCites Journal Citation Reports 2016. https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action (accessed December 14, 2016).

[21] Guertin KA, Freedman ND, Loftfield E, Graubard BI, Caporaso NE, Sinha R. Coffee consumption and incidence of lung cancer in the NIH-AARP Diet and Health Study. Int J Epidemiol 2016;45:929–39. doi:10.1093/ije/dyv104.

[22] Song M, Hu FB, Spiegelman D, Chan AT, Wu K, Ogino S, et al. Long-term status and

change of body fat distribution, and risk of colorectal cancer: a prospective cohort study. Int J Epidemiol 2016;45:871–83. doi:10.1093/ije/dyv177.

[23] Pedersen SB, Langsted A, Nordestgaard BG. Nonfasting mild-to-moderate hypertriglyceridemia and risk of acute pancreatitis. JAMA Intern Med 2016;176:1834–42. doi:10.1001/jamainternmed.2016.6875.

[24] Knuiman MW, Divitini ML, Buzas JS, Fitzgerald PEB. Adjustment for regression dilution in epidemiological regression analyses. Ann Epidemiol 1998;8:56–63. doi:10.1016/S1047-2797(97)00107-5.

[25] Wallace ME, Grantz KL, Liu D, Zhu Y, Kim SS, Mendola P. Exposure to ambient air pollution and premature rupture of membranes. Am J Epidemiol 2016;183:1114–21. doi:10.1093/aje/kwv284.

[26] Pongiglione B, De Stavola BL, Kuper H, Ploubidis GB. Disability and all-cause mortality in the older population: evidence from the English Longitudinal Study of Ageing. Eur J Epidemiol 2016;31:735–46. doi:10.1007/s10654-016-0160-8.

[27] Mitchell KS, Porter B, Boyko EJ, Field AE. Longitudinal associations among posttraumatic stress disorder, disordered eating, and weight gain in military men and women. Am J Epidemiol 2016;184:33–47. doi:10.1093/aje/kwv291.

[28] Leslie WD, Majumdar SR, Morin SN, Lix LM. Change in bone mineral density is an indicator of treatment-related antifracture effect in routine clinical practice: a registry-based cohort study. Ann Intern Med 2016;165:465–72. doi:10.7326/M15-2937.

[29] Turkiewicz A, Neogi T, Björk J, Peat G, Englund M. All-cause mortality in knee and hip osteoarthritis and rheumatoid arthritis. Epidemiology 2016;27:479–85. doi:10.1097/EDE.0000000000000477.

[30] Clausen TD, Bergholt T, Eriksson F, Rasmussen S, Keiding N, Løkkegaard EC. Prelabor cesarean section and risk of childhood type 1 diabetes: a nationwide register-

based cohort study. Epidemiology 2016;27:547–55. doi:10.1097/EDE.0000000000000488.

[31] Auger N, Fraser WD, Smargiassi A, Bilodeau-Bertrand M, Kosatsky T. Elevated outdoor temperatures and risk of stillbirth. Int J Epidemiol 2016;46:200–8. doi:10.1093/ije/dyw077.

[32] Dawson AL, Tinker SC, Jamieson DJ, Hobbs CA, Berry RJ, Rasmussen SA, et al. Twinning and major birth defects, National Birth Defects Prevention Study, 1997-2007. J Epidemiol Community Health 2016;70:1114–21. doi:10.1136/jech-2015-206302.

[33] Svanes C, Koplin J, Skulstad SM, Johannessen A, Bertelsen RJ, Benediktsdottir B, et al. Father's environment before conception and asthma risk in his children: a multi-generation analysis of the Respiratory Health In Northern Europe study. Int J Epidemiol 2016;46:235–45. doi:10.1093/ije/dyw151.

[34] Gerber JS, Bryan M, Ross RK, Daymont C, Parks EP, Localio AR, et al. Antibiotic exposure during the first 6 months of life and weight gain during childhood. JAMA 2016;315:1258–65. doi:10.1001/jama.2016.2395.

[35] Menvielle G, Franck J, Radoi L, Sanchez M, Févotte J, Guizard AV, et al. Quantifying the mediating effects of smoking and occupational exposures in the relation between education and lung cancer: the ICARE study. Eur J Epidemiol 2016;31:1213–21. doi:10.1007/s10654-016-0182-2.

[36] Graham DJ, Reichman ME, Wernecke M, Hsueh Y-H, Izem R, Southworth MR, et al. Stroke, bleeding, and mortality risks in elderly medicare beneficiaries treated with dabigatran or rivaroxaban for nonvalvular atrial fibrillation. JAMA Intern Med 2016;176:1662–71. doi:10.1001/jamainternmed.2016.5954.

[37] Martinez C, Suissa S, Rietbrock S, Katholing A, Freedman B, Cohen AT, et al.

Testosterone treatment and risk of venous thromboembolism: population based case-control study. BMJ 2016;355:1–9. doi:10.1136/bmj.i5968.

[38] Upson K, Harmon QE, Laughlin-Tommaso SK, Umbach DM, Baird DD. Soy-based infant formula feeding and heavy menstrual bleeding among young African American omen. Epidemiology 2016;27:716–25. doi:10.1097/EDE.0000000000000508.

[39] Bodnar LM, Pugh SJ, Lash TL, Hutcheon JA, Himes KP, Parisi SM, et al. Low gestational weight gain and risk of adverse perinatal outcomes in obese and severely obese women. Epidemiology 2016;27:894–902. doi:10.1097/EDE.0000000000000535.

[40] R Core Team. R: a language and environment for statistical computing 2014.

[41] Lederer W, Küchenhoff H. simex: SIMEX- and MCSIMEX-Algorithm for measurement error models 2013.

[42] Rosseel Y. lavaan : an R package for structural equation modeling. J Stat Softw 2012;48:1–20.

[43] Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? Am J Epidemiol 1990;132:373–5.

[44] Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: Expectations vs observations. Int J Epidemiol 2005;34:680–7. doi:10.1093/ije/dyi060.

[45] Loken E, Gelman A. Measurement error and the replication crisis. Science (80-) 2017;355:584–5. doi:10.1126/science.aal3618.

[46] Rutjes A, Reitsma J, Coomarasamy A, Khan K, Bossuyt P. Evaluation of diagnostic tests when there is no gold standard- a review of methods. Health Technol Assess (Rockv) 2007;11:1–4. doi:06/90/23 [pii].

[47] Collins J, Huynh M. Estimation of diagnostic test accuracy without full verification: A

review of latent class methods. Stat Med 2014;33:4141–69. doi:10.1002/sim.6218.

[48] van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KGM, de Groot JAH. Latent Class Models in Diagnostic Studies When There is No Reference Standard--A Systematic Review. Am J Epidemiol 2014;179:423–31. doi:10.1093/aje/kwt286.

APPENDIX A

PubMed Search String

“(("N Engl J Med"[Journal] OR "Lancet"[Journal] OR "JAMA"[Journal] OR "BMJ"[Journal] OR "Ann Intern Med"[Journal] OR "JAMA Intern Med"[Journal] OR "Int J Epidemiol"[Journal] OR "Eur J Epidemiol"[Journal] OR "Epidemiology"[Journal] OR "Am J Epidemiol"[Journal] OR "J Clin Epidemiol"[Journal] OR "J Epidemiol Community Health"[Journal]) AND ("2016/01/01"[PDAT] : "2016/12/31"[PDAT]) NOT (Addresses[ptyp] OR Autobiography[ptyp] OR Bibliography[ptyp] OR Biography[ptyp] OR Clinical Conference[ptyp] OR Comment[sb] OR Congresses[ptyp] OR Consensus Development Conference[ptyp] OR Consensus Development Conference, NIH[ptyp] OR Dictionary[ptyp] OR Directory[ptyp] OR Editorial[ptyp] OR Festschrift[ptyp] OR Interactive Tutorial[ptyp] OR Introductory Journal Article[ptyp] OR Lectures[ptyp] OR Legal Cases[ptyp] OR Legislation[ptyp] OR Meta-Analysis[ptyp] OR News[ptyp] OR Newspaper Article[ptyp] OR Patient Education Handout[ptyp] OR Personal Narratives[ptyp] OR Portraits[ptyp] OR Research Support, American Recovery and Reinvestment Act[ptyp] OR Research Support, N I H, Extramural[ptyp] OR Research Support, N I H, Intramural[ptyp] OR Research Support, Non U S Gov't[ptyp] OR Research Support, U S Gov't, Non P H S[ptyp] OR Research Support, U S Gov't, P H S[ptyp] OR Research Support, U.S. Government[ptyp] OR Retracted Publication[sb] OR Retraction of Publication[sb] OR Review[ptyp] OR Scientific Integrity Review[ptyp] OR systematic[sb] OR Video-Audio Media[ptyp] OR Webcasts[ptyp] OR "published erratum"[Publication Type] OR "case reports"[Publication Type] OR "historical article"[Publication Type] OR "letter"[Publication Type])) AND hasabstract[text]”

APPENDIX B

Adobe Reader XI Pro Full-Text Search String (With Options "Boolean" and "Stemming" Enabled)

"measurement error OR error measure OR measured with error OR error in measure OR mismeasure OR insensitive measurement OR unspecific measurement OR information bias OR misclassify OR misclassification OR classification error OR attenuate OR residual confounding OR miscode OR coding mistake OR deattenuate OR error in assessment OR bias measurement OR errant measure OR measure errantly OR measure erroneous OR erroneous measure OR self-report OR self-reported"

APPENDIX C

Data Extraction Form Used for 247 Publications That Reported on Measurement Error

Basic information

1. Accession # [open text field]
2. Journal [drop down menu with all relevant journals]
3. First Author Last name [open text field]
4. (E)pub date [date field DD-MM-YY]
5. Extractor [drop down menu with author initials]
6. Date of extraction [date field DD-MM-YY]

General Measurement Error Information

7. Are measurement error or related terms identified in the article? [y/n]
8. Is at least 1 hit **relevant** to measurement error? [y/n]
9. To which variables do the **relevant** measurement error statements pertain? [multiple answers: exposure(s), confounder(s), outcome(s), unclear]
10. In which sections of the article are **relevant** measurement error statements made? [multiple answers: introduction, methods, results, conclusion/discussion]
    b. other (specify) [open text field]
11. Any attempt to correct for measurement error in this study? [y/n]
32. Doubt if the **relevant** measurement error is in the exposure or confounder? [y/n]
33. Is measurement error discussed with respect to previous research? [y/n]
34. Are measurement error statements made with respect to the prevention of measurement error in the current study? [y/n]

Studied Exposures

12. Was/were the exposure(s) randomized (i.e. is the study a randomized trial) [y/n]
13. Is the study a prediction study? [y/n]
    (***IF YES, CONSIDER Q17 - Q22 AS PERTAINING TO THE PREDICTORS)***
14. Type of exposure(s) [multiple answers: lifestyle/dietary, biological processes, medical/ drug intervention, demographic factors, infrastructure]
    a. other(specify) [open text field]
15. Were any of the measurement error statements pertaining to the exposure(s)? [dropdown: Yes, No, Unclear, Not Applicable] (***IF NO, Q19 – Q22 are to be filled in with NA)***
16. In which section of the article was it discussed [multiple answers: introduction, methods, results, conclusion/discussion, Not applicable]
    a. Other [open text field]
17. Was there ME in the exposure? Direction of the anticipated effect of ME on studied relations? [dropdown: Yes, underestimation of true effect; Yes, overestimation of true effect; Yes, no/negligible effect; Yes, Unclear/Not mentioned; No; Not Applicable]
18. Was any statement made about the type of ME? [multiple answers: Differentiality , Error model, Not mentioned, Not Applicable]
    a. Other [open text field]
19. Was any attempt made to correct for the ME? [dropdown: Yes, No, Unclear, Not Applicable]
20. Notes/Citations of exposure ME and its effect [open text field]

Studied Confounders/ Predictors

21. Type of confounder(s) [multiple answers: lifestyle/dietary, biological processes, medical/drug intervention, demographic factors, infrastructure]

a. Other [open text field]

22. Was measurement error of confounder(s) mentioned? [dropdown: Yes, No, Unclear, Not Applicable (*If no confounders in the study)*] ***IF NO, Q26 – Q29 ARE NA***
23. In which section of the article was it discussed [multiple answers: introduction, methods, results, conclusion/discussion, Not applicable]
    a. Other [open text field]
24. Was there ME in the confounder(s)? Direction of the anticipated effect of ME on studied relations? [dropdown: Yes, underestimation of true effect; Yes, overestimation of true effect; Yes, no/negligible effect; Yes, Unclear/Not mentioned; No; Not Applicable]
25. Was any statement made about the type of ME? [multiple answers: Differentiality , Error model, Not mentioned, Not Applicable]
    a. Other [open text field]
26. Was any attempt made to correct for the ME? [dropdown: Yes, No, Unclear, Not Applicable]
27. Notes/Citations of confounder ME and its effect [open text field]

Studied Outcome

28. Type of outcome(s) [multiple answers: mortality, CVD, cancer, infections, hospitalization]
    a. Other [open text field]
29. Notes on type of outcome [open text field]

Next Steps

30. Include article for full extraction [dropdown menu: Yes, No, Maybe, Fully disregard (false positive)]

31. Any other general comments [open text field]

Revise form together? [y/n]