

Accepted Manuscript

Measurement error is often neglected in medical literature: a systematic review

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



PII: S0895-4356(17)31083-1

DOI: [10.1016/j.jclinepi.2018.02.023](https://doi.org/10.1016/j.jclinepi.2018.02.023)

Reference: JCE 9612

To appear in: *Journal of Clinical Epidemiology*

Received Date: 25 September 2017

Revised Date: 1 February 2018

Accepted Date: 28 February 2018

Please cite this article as: Brakenhoff TB, Mitroiu M, Keogh RH, Moons KGM, Groenwold RHH, van Smeden M, Measurement error is often neglected in medical literature: a systematic review, *Journal of Clinical Epidemiology* (2018), doi: 10.1016/j.jclinepi.2018.02.023.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 **Measurement error is often neglected in medical literature: a systematic review**

2

3 Timo B. Brakenhoff, MSc¹, Marian Mitroiu, MSc¹, Ruth H Keogh, PhD², Karel G.M. Moons, PhD¹,

4 Rolf H.H. Groenwold, MD, PhD¹, Maarten van Smeden, PhD¹

5

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

7 Netherlands

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

9

10 *Corresponding author:*

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

12 Julius Center for Health Sciences and Primary Care

13 University Medical Center Utrecht

14 PO Box 85500, 3508 GA Utrecht, the Netherlands

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

16

17

18

19

20

21 ABSTRACT

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

37

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

39

40

41 WHAT'S NEW

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

56

57 1. Introduction

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

65
66 It is generally well accepted that measurement error and classification error (hereinafter
67 collectively referred to as measurement error) in either the dependent variable (hereinafter
68 *outcome*) or independent explanatory variables (hereinafter *covariates*; e.g. exposure and
69 confounder variables) can introduce bias and imprecision to estimates of covariate-outcome
70 relations. Among others, several textbooks [3–6], methodological reviews [7,8] and a tool-kit
71 [9], have demonstrated how to examine, quantify, and correct for measurement error in a
72 variety of settings encountered in epidemiology. Most of this work has been focused on
73 measurement error in covariates given its conceived greater impact on studied relations than
74 measurement error in the outcome [4]. Despite these resources, it is suspected that the
75 attention it receives in applied medical and epidemiological studies is insufficient [10,11].

76
77 Over a decade ago, a review of 57 randomly selected publications from three high ranking
78 epidemiology journals reported that 61% of the reviewed publications recognized the
79 potential influence of measurement error, but only 28% made a qualitative assessment of its
80 impact on their results, and only one quantified its potential impact on results [12]. In light of
81 the increasing prevalence of measurement error in medical and epidemiological research and

82 increasing availability of methods and software to account for measurement error, a new and
83 more comprehensive investigation into current practice is necessary.

84

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

89

90 **2. Measurement error**

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

100

101 In the context of multivariable statistical models, such as regression models, measurement
102 error can be present in the outcome and/or covariates. We focus on error in covariates. In their
103 seminal text-book, Carroll et al. [5] describe the effect of measurement error in covariates as a
104 “triple whammy”: covariate-outcome relationships can be biased, power to detect clinically
105 meaningful relationships is diminished, and features of the data can be masked. Whether bias
106 is present, and if so its direction and magnitude, depend on the form of the measurement

107 error. It is therefore important to quantify any bias due to measurement error and to obtain
108 corrected estimates where possible. Three important considerations in this process are:
109 identification of the variables of interest that are measured with error, what type of
110 measurement error is present, and what additional information is available to help characterize
111 the error.

112

113 *2.1 Types of measurement error and their effects*

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

117

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

131 this bias is thus unpredictable and this holds for different regression models of interest in
132 epidemiology, including logistic, Cox and linear regression models [5].

133

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

144

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

154

155 For categorical variables, measurement error is commonly referred to as *misclassification*.
156 Misclassification can be summarized using sensitivity and specificity when the variable is
157 binary. In the situation where a single binary exposure is related to an outcome, random non-
158 differential misclassification present in the exposure will result in attenuation of this
159 exposure-outcome relation [1]. However, when the exposure has more than two categories,
160 when the exposure is subject to systematic or differential misclassification, or when
161 confounders measured with error are added to the analysis model, it is once more difficult to
162 predict in which direction the estimate of the true exposure-outcome relation will be biased
163 [4].

164

165 *2.2 Measurement error correction methods*

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

170

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

177

178 Cook and Stefanski proposed the simulation-extrapolation (SIMEX) method [16]. This
179 method works via a two-step procedure. First, data are simulated by adding additional error of

180 different magnitudes to the observed exposure measurements; the simulated data sets are used
181 to estimate the effect of this additional error on the exposure-outcome relation. As a second
182 step, the estimate of the exposure-outcome relation is extrapolated back to the situation where
183 there is no measurement error using an extrapolation model which relates the estimated
184 exposure-outcome association parameter to the degree of measurement error. Like regression
185 calibration, this method requires information about the amount of measurement error
186 (variance) in the observed variable. SIMEX as described above assumes non-differential
187 classical error, yet has also been extended to deal with misclassified categorical variables
188 [17].

189
190 Alternatively, a large range of so-called latent variable models have been suggested to
191 account for measurement error during analysis. Latent variable models generally rely on
192 replicate measurements of error-prone measures to estimate a latent variable to represent the
193 true error-free variable [18]. This latent variable can replace the observed error-prone variable
194 in the exposure-outcome analysis or can be modelled directly in the exposure-outcome model,
195 for instance, using Structural Equation Modeling [18,19].

196
197 We acknowledge that it can be very challenging to determine the structure and amount of
198 measurement error due to the plethora of underlying (unobserved) factors that may influence
199 it. While further guidance is required on how to assess the amount and type of measurement
200 error in practice, it can generally be recommended to collect additional data, whenever
201 feasible, either in a subset of the study sample or possibly in an external validation sample, to
202 compare observations on a covariate that is (suspected of being) measured with error and an
203 error free representation of that covariate (if such a 'gold standard' exists). This information
204 can subsequently be used to study measurement error structures, amount of measurement

205 error, and to inform measurement error correction methods (e.g. regression calibration or
206 SIMEX, among others), which allow for a measurement error corrected analysis on the whole
207 study sample. Alternatively, when available, repeated measurements of a covariate measured
208 with error can be used to quantify measurement error variance and allow for measurement
209 error corrected analyses.

210

211 *2.3 Availability of additional information for measurement error corrections*

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

215

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

222

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

228

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

231

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

241

242 3. Methods

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

249
250 Using the Thomson Reuters InCites rankings of 2015 [20], the 6 highest-ranking journals in
251 the categories “General & Internal Medicine” (New England Journal of Medicine, Lancet,
252 JAMA, BMJ, Annals of Internal Medicine and JAMA Internal Medicine) and
253 “Epidemiology” (International Journal of Epidemiology, European Journal of Epidemiology,
254 Epidemiology, American Journal of Epidemiology, Journal of Clinical Epidemiology, Journal
255 of Epidemiology and Community Health) were identified. The journal Epidemiology Review
256 was excluded as it is an annual journal. All publications of the above-mentioned journals from
257 the period 01/01/2016 to 31/12/2016 were identified using PubMed (see search string in
258 Appendix A).

259
260 Title and abstracts were screened by one reviewer (TB). Publications that were not original
261 research (e.g. brief reports, essays, cohort profiles, and guidance papers) were excluded. Also
262 excluded were: methodological research, review and meta-analysis research, qualitative
263 research, policy oriented studies, descriptive studies, studies that analyzed data on an
264 aggregated level, and publications that did not assess individual health related exposures and
265 outcomes.

266

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

278
279 The included publications were reviewed independently by two readers (TB and MM) using a
280 standardized data extraction form (see Appendix C). This form was pilot tested by four
281 researchers (TB, MS, RG, MM). Disagreements were discussed until consensus was reached.
282 The elements extracted included: design of data collection, study characteristics, clinical
283 domain, characterization of variable(s) subject to measurement error (exposure/confounder),
284 sections of the article where measurement error was mentioned
285 (abstract/introduction/methods/results/discussion), reporting of possible effects of
286 measurement error on study results (direction and magnitude of effect), reporting of the
287 assumed type of error, reporting of methods that investigated the impact of, or attempted to
288 correct for, measurement error in exposure or confounder variables.

289
290 Articles that reported impact of measurement error or corrections for measurement error were
291 included for additional review by four readers (TB, MS, RG, MM). For these publications,

292 data were extracted from the main document and the supplementary materials. The methods
293 used were characterized, alongside how this was reported and the type of additional
294 information used.

295

296 **4. Results**

297 Figure 1 depicts the number of included papers at each step of the review process. Of the
298 1178 articles found in PubMed, 565 (337 from Epidemiology journals and 228 from General
299 & Internal Medicine journals) were judged as original research satisfying our inclusion
300 criteria. Of these, 247 (44%) directly addressed measurement error in some form.
301 Characteristics of these included studies are found in Table 1. Eighteen of these publications
302 (3% of the 565) investigated the possible impact of, or corrected for, measurement error.
303 Thirteen of these eighteen publications were from Epidemiology journals (4% of the 337
304 Epidemiology publications) and the remaining five were from General & Internal Medicine
305 Journals (2% of the 228 General & Internal Medicine publications). Table 2 shows from
306 which journals the publications that directly addressed measurement error originated.

307

308

309 >> **insert Fig. 1** Flow Diagram Detailing the Systematic Review Process<<

310

311 >> **insert Table 1** General Characteristics of the 247 Publications That Explicitly Report on
312 Measurement Error (ME) in Some Form.<<<

313

314 ME = Measurement error

315 ^a 174 (70%) publications considered ME **only** in the discussion section

316 ^b Mentions made of ME pertained to previously published research and not to the study presented in the
317 published paper.

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

319

320

321

322

323

324

325

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

328
329 ME=Measurement error

330
331
332

333 *4.1 Measurement error in main exposure variables*

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

344

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

352

353 *4.2 Measurement error in confounder variables*

354 Of the 44 publications that reported on measurement error in the confounders, 29 (66%)
355 reported the presence of measurement error without mentioning (or were unclear about) its
356 possible effect on the studied relations, six (14%) reported that the measurement error in the
357 confounder did or could have led to underestimation of the relation between the main
358 exposure and the outcome, and four (9%) reported that measurement error in the confounder
359 was anticipated to have no or only a negligible effect on the main exposure–outcome relation.
360 None of the publications reported on possible overestimation of the main exposure-outcome
361 relation due to confounders measured with error. Five (11%) publications explicitly reported
362 that their confounder variable(s) were measured *without* error.
363 Six (14%) of the 44 publications made general statements about the structure of the
364 measurement error. One discussed the assumed error model.

365

366 *4.3 Measurement error impact and correction*

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

370

371

372 >> **Insert Table 3** Characteristics of the 18 Publications That Reported on Investigation of or
373 Correction for Measurement Error (ME).<<

374

375 ME=Measurement error

376 *Methods designed specifically for a field of applied research

377

378 Seven publications (39%) of the 18, applied measurement error correction methods. Two
379 publications used regression calibration, relying on internal validation data. One of these [21]
380 used additional data gathered for a subset of participants to account for measurement error in
381 the exposure (daily coffee intake). The other [22] corrected for measurement error in several

382 anthropomorphic measurements using data from earlier validation studies conducted within
383 the same cohort. One publication [23] used a non-parametric method [24] to correct for
384 underestimation of the exposure-outcome relation because of assumed random measurement
385 error in the exposure (plasma triglycerides values at baseline). Another publication [25] used
386 external observed air quality monitoring data to correct their estimates of individual air
387 pollutant exposure. Two publications used factor analysis to define a latent exposure. One
388 [26] implemented a latent variable model to determine each individual's disability score using
389 many different items of a conceptual framework for describing functioning and disability.
390 This score was then used in a regression analysis. In another [27] the factor analysis was
391 embedded in a structural equation model where latent PTSD status was estimated from
392 multiple clusters of symptoms suggestive of PTSD. Finally, Leslie et al. [28] used an ad-hoc
393 approach, coined 'least significant change', to take into account inherent instrument
394 measurement error when ascertaining exposure status (absolute bone mineral density
395 difference).

396
397 The remaining 11 (61%) of the 18 publications investigated the impact of measurement error
398 on the exposure-outcome relation using sensitivity analyses. In five publications [29–33], an
399 assumption was made about the amount of possible measurement error and its effect on the
400 exposure-outcome relation was quantified. Often this was achieved by looking at a subgroup
401 of the original sample for which the mismeasured variable of interest was assumed to be
402 measured with less or no error. Four publications [34–37] looked at multiple scenarios in
403 which they assumed different amounts of measurement error. The remaining two publications
404 [38,39] performed a probabilistic sensitivity analysis. All authors reported that the results of
405 the sensitivity analyses were either similar to those of the conventional analyses or did not

406 influence their conclusions. No study investigated the impact of measurement error on their
407 results using an external dataset.

408

409

410

411

412

ACCEPTED MANUSCRIPT

413 5. Discussion

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

420

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

436

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

442
443 As compared to the abundance of qualitative statements made about the presence of
444 measurement error, we found formal measurement error evaluations to be surprisingly rare.
445 About 4% of the papers that made a qualitative statement about measurement error quantified
446 its impact using sensitivity analyses. Only 2% used formal measurement error correction
447 methods. Several reasons for this low prevalence can be postulated. In practice it can be very
448 challenging to properly assess the structure and amount of measurement error. Obviously,
449 determining a strategy to account for measurement error in the analysis is then very difficult.
450 But even when a suitable strategy can be determined and data are available to implement the
451 strategy, there may still be lack of familiarity with these methods and available software
452 among applied researchers, medical readers and journal editors, which may frustrate the
453 adoption of these methods in the medical literature. For example, statistical software such as
454 R [40] can be used to implement regression calibration (see supplementary material of [9]),
455 SIMEX [41] and latent variable modeling [42]. There also seems to be a lack of educational
456 materials and courses that provide guidance for practicing researchers, peer-reviewers and
457 editors on how to use, assess and interpret results from measurement error correction
458 methods.

459
460 A need for better understanding of measurement error in medical and epidemiologic research
461 is further supported by a noticeably high incidence (about one third of those that discussed

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

472
473 Of the 18 publications that investigated or corrected for measurement error, most manuscripts
474 reported both the original ('naïve') and the measurement error corrected results.
475 Unfortunately, descriptions of the used methods were often not provided. Indeed, half of the
476 publications that performed sensitivity analyses reported the results using only a single line in
477 the results section claiming similarity of results to the main analysis (e.g., [36]). A similar
478 proportion of these publications also only investigated one possible measurement error
479 scenario.

480
481 Our review has some limitations. It cannot be ruled out that our full-text search strategy may
482 have missed papers that mentioned measurement error. Although our search string covered a
483 broad range of terminology related to measurement error, papers using atypical terms may
484 have been overlooked. This might have led to an underestimation of the number of
485 publications that discussed measurement error. This limitation is unlikely to have a substantial
486 impact on the estimated percentages and conclusions, given that the intention was to give a

487 general impression of current practice of measurement error reporting. Second, in our review
488 we ignored measurement error issues related to the outcome variable. While measurement
489 error in outcome variables is often assumed to pose less problems than measurement error in
490 covariates [4], we acknowledge that this choice limits our findings. Finally, there are
491 measurement errors that influence analyses that do not strictly fall in the multivariable
492 (exposure – outcome) classification. Specifically, diagnostic test accuracy studies often suffer
493 from measurement error in the disease verification procedure, a problem known as “absence
494 of gold standard”, and were outside the scope of this review. Reviews of methods [46,47] and
495 the use of methods [48] to account for disease verification problems are found elsewhere.

496

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

505

506 In conclusion, we found that measurement error is often discussed in high impact medical and
507 epidemiologic literature. However, only a small portion proceeds to investigate or correct for
508 measurement error. Renewed efforts are required to raise awareness among applied
509 researchers that measurement error can have a large impact on estimated exposure-outcome
510 relations and that tools are available to quantify this impact. More guidance and tutorials seem
511 necessary to assist the applied researchers with the assessment of the type and amount of

512 measurement error as well as the steps that can subsequently be taken to minimize its impact
513 on the studied relations. Given the unpredictable nature of the impact of measurement error
514 on the studied results, we advise authors to report on the potential presence of measurement
515 error in recorded variables but exercise restraint when speculating about the magnitude and
516 direction of its impact unless the appropriate analysis steps are taken to substantiate such
517 claims. Also, we recommend authors to make more use of available correction methods and
518 probabilistic sensitivity analyses to correct analyses for variables that were measured with
519 error. Given the increasing use of data not originally intended for medical or epidemiological
520 research, we anticipate that the use and understanding of measurement error analyses and
521 corrections will become increasingly important in the near future.

522 FUNDING

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

525 CONFLICT OF INTEREST

526 Conflicts of interest: none

527

ACCEPTED MANUSCRIPT

528

529 REFERENCES

- 530 [1] Rothman KJ, Greenland S, Lash TL, editors. *Modern Epidemiology*. 3rd ed.
531 Philadelphia, PA, USA: Lippincott Williams & Wilkins; 2008.
- 532 [2] Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and
533 Clinical Medicine. *N Engl J Med* 2016;375:1216–9. doi:10.1002/aur.1474.Replication.
- 534 [3] Fuller WA. *Measurement Error Models*. John Wiley & Sons; 1987.
- 535 [4] Gustafson P. *Measurement Error and Misclassification in Statistics and Epidemiology:
536 Impacts and Bayesian Adjustments*. Boca Raton, United States: Chapman and
537 Hall/CRC; 2004.
- 538 [5] Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement error in nonlinear
539 models: a modern perspective*. 2nd ed. Chapman & Hall /CRC Press; 2006.
- 540 [6] Buonaccorsi J. *Measurement Error: Models, Methods and Applications*. CRC Press;
541 2010.
- 542 [7] Stefanski LA. *Measurement Error Models*. *J Am Stat Assoc* 2000;95:1353–8.
- 543 [8] Guolo A. Robust techniques for measurement error correction: a review. *Stat Methods
544 Med Res* 2008;17:555–80. doi:10.1177/0962280207081318.
- 545 [9] Keogh R, White I. A toolkit for measurement error correction, with a focus on
546 nutritional epidemiology. *Stat Med* 2014;33:2137–55. doi:10.1002/sim.6095.
- 547 [10] Buzas JS, Stefanski LA, Tosteson TD. Measurement Error. In: Ahrens W, Pigeot I,
548 editors. *Handb. Epidemiol.*, 2014, p. 1241–82. doi:10.1007/978-0-387-09834-0.
- 549 [11] Blackwell M, Honaker J, King G. A Unified Approach to Measurement Error and
550 Missing Data: Overview and Applications. *Sociol Methods Res* 2015:1–39.
551 doi:10.1177/0049124115589052.
- 552 [12] Jurek AM, Maldonado G, Greenland S, Church TR. Exposure-measurement error is

- 553 frequently ignored when interpreting epidemiologic study results. *Eur J Epidemiol*
554 2006;21:871–6. doi:10.1007/s10654-006-9083-0.
- 555 [13] Brakenhoff TB, van Smeden M, Visseren FL, Groenwold RHH. Random measurement
556 error: why worry? An example of cardiovascular risk factors. *PLoS One* 2018;In Press.
- 557 [14] Ahrens W, Pigeot I, editors. *Handbook of Epidemiology*. 2nd ed. New York, USA:
558 Springer-Verlag New York; 2014.
- 559 [15] Rosner B, Willett W, Spiegelman D. Correction of logistic regression relative risk
560 estimates and confidence intervals for systematic within-person measurement error.
561 *Stat Med* 1989;8:1051–69.
- 562 [16] Cook J, Stefanski L. Simulation-extrapolation estimation in parametric measurement
563 error models. *J Am Stat Assoc* 1994;89:1314–28. doi:10.2307/2290994.
- 564 [17] Küchenhoff H, Mwalili SM, Lesaffre E. A general method for dealing with
565 misclassification in regression: The misclassification SIMEX. *Biometrics* 2006;62:85–
566 96. doi:10.1111/j.1541-0420.2005.00396.x.
- 567 [18] Skrondal A, Rabe-Hesketh S. *Generalized latent variable modeling: Multilevel,
568 longitudinal, and structural equation models*. Crc Press; 2004.
- 569 [19] Kline RB. *Principles and practice of structural equation modeling*. Guilford
570 publications; 2015.
- 571 [20] Thomson Reuters. *InCites Journal Citation Reports 2016*.
572 <https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action> (accessed
573 December 14, 2016).
- 574 [21] Guertin KA, Freedman ND, Loftfield E, Graubard BI, Caporaso NE, Sinha R. Coffee
575 consumption and incidence of lung cancer in the NIH-AARP Diet and Health Study.
576 *Int J Epidemiol* 2016;45:929–39. doi:10.1093/ije/dyv104.
- 577 [22] Song M, Hu FB, Spiegelman D, Chan AT, Wu K, Ogino S, et al. Long-term status and

- 578 change of body fat distribution, and risk of colorectal cancer: a prospective cohort
579 study. *Int J Epidemiol* 2016;45:871–83. doi:10.1093/ije/dyv177.
- 580 [23] Pedersen SB, Langsted A, Nordestgaard BG. Nonfasting mild-to-moderate
581 hypertriglyceridemia and risk of acute pancreatitis. *JAMA Intern Med* 2016;176:1834–
582 42. doi:10.1001/jamainternmed.2016.6875.
- 583 [24] Knuiman MW, Divitini ML, Buzas JS, Fitzgerald PEB. Adjustment for regression
584 dilution in epidemiological regression analyses. *Ann Epidemiol* 1998;8:56–63.
585 doi:10.1016/S1047-2797(97)00107-5.
- 586 [25] Wallace ME, Grantz KL, Liu D, Zhu Y, Kim SS, Mendola P. Exposure to ambient air
587 pollution and premature rupture of membranes. *Am J Epidemiol* 2016;183:1114–21.
588 doi:10.1093/aje/kwv284.
- 589 [26] Pongiglione B, De Stavola BL, Kuper H, Ploubidis GB. Disability and all-cause
590 mortality in the older population: evidence from the English Longitudinal Study of
591 Ageing. *Eur J Epidemiol* 2016;31:735–46. doi:10.1007/s10654-016-0160-8.
- 592 [27] Mitchell KS, Porter B, Boyko EJ, Field AE. Longitudinal associations among
593 posttraumatic stress disorder, disordered eating, and weight gain in military men and
594 women. *Am J Epidemiol* 2016;184:33–47. doi:10.1093/aje/kwv291.
- 595 [28] Leslie WD, Majumdar SR, Morin SN, Lix LM. Change in bone mineral density is an
596 indicator of treatment-related antifracture effect in routine clinical practice: a registry-
597 based cohort study. *Ann Intern Med* 2016;165:465–72. doi:10.7326/M15-2937.
- 598 [29] Turkiewicz A, Neogi T, Björk J, Peat G, Englund M. All-cause mortality in knee and
599 hip osteoarthritis and rheumatoid arthritis. *Epidemiology* 2016;27:479–85.
600 doi:10.1097/EDE.0000000000000477.
- 601 [30] Clausen TD, Bergholt T, Eriksson F, Rasmussen S, Keiding N, Løkkegaard EC.
602 Prelabor cesarean section and risk of childhood type 1 diabetes: a nationwide register-

- 603 based cohort study. *Epidemiology* 2016;27:547–55.
604 doi:10.1097/EDE.0000000000000488.
- 605 [31] Auger N, Fraser WD, Smargiassi A, Bilodeau-Bertrand M, Kosatsky T. Elevated
606 outdoor temperatures and risk of stillbirth. *Int J Epidemiol* 2016;46:200–8.
607 doi:10.1093/ije/dyw077.
- 608 [32] Dawson AL, Tinker SC, Jamieson DJ, Hobbs CA, Berry RJ, Rasmussen SA, et al.
609 Twinning and major birth defects, National Birth Defects Prevention Study, 1997-
610 2007. *J Epidemiol Community Health* 2016;70:1114–21. doi:10.1136/jech-2015-
611 206302.
- 612 [33] Svanes C, Koplein J, Skulstad SM, Johannessen A, Bertelsen RJ, Benediktsdottir B, et
613 al. Father’s environment before conception and asthma risk in his children: a multi-
614 generation analysis of the Respiratory Health In Northern Europe study. *Int J*
615 *Epidemiol* 2016;46:235–45. doi:10.1093/ije/dyw151.
- 616 [34] Gerber JS, Bryan M, Ross RK, Daymont C, Parks EP, Localio AR, et al. Antibiotic
617 exposure during the first 6 months of life and weight gain during childhood. *JAMA*
618 2016;315:1258–65. doi:10.1001/jama.2016.2395.
- 619 [35] Menvielle G, Franck J, Radoi L, Sanchez M, Févotte J, Guizard AV, et al. Quantifying
620 the mediating effects of smoking and occupational exposures in the relation between
621 education and lung cancer: the ICARE study. *Eur J Epidemiol* 2016;31:1213–21.
622 doi:10.1007/s10654-016-0182-2.
- 623 [36] Graham DJ, Reichman ME, Wernecke M, Hsueh Y-H, Izem R, Southworth MR, et al.
624 Stroke, bleeding, and mortality risks in elderly medicare beneficiaries treated with
625 dabigatran or rivaroxaban for nonvalvular atrial fibrillation. *JAMA Intern Med*
626 2016;176:1662–71. doi:10.1001/jamainternmed.2016.5954.
- 627 [37] Martinez C, Suissa S, Rietbrock S, Katholing A, Freedman B, Cohen AT, et al.

- 628 Testosterone treatment and risk of venous thromboembolism: population based case-
629 control study. *BMJ* 2016;355:1–9. doi:10.1136/bmj.i5968.
- 630 [38] Upson K, Harmon QE, Laughlin-Tommaso SK, Umbach DM, Baird DD. Soy-based
631 infant formula feeding and heavy menstrual bleeding among young African American
632 women. *Epidemiology* 2016;27:716–25. doi:10.1097/EDE.0000000000000508.
- 633 [39] Bodnar LM, Pugh SJ, Lash TL, Hutcheon JA, Himes KP, Parisi SM, et al. Low
634 gestational weight gain and risk of adverse perinatal outcomes in obese and severely
635 obese women. *Epidemiology* 2016;27:894–902. doi:10.1097/EDE.0000000000000535.
- 636 [40] R Core Team. *R: a language and environment for statistical computing* 2014.
- 637 [41] Lederer W, Küchenhoff H. *simex: SIMEX- and MCSIMEX-Algorithm for*
638 *measurement error models* 2013.
- 639 [42] Rosseel Y. *lavaan : an R package for structural equation modeling*. *J Stat Softw*
640 2012;48:1–20.
- 641 [43] Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of
642 exposure always bias a true effect toward the null value? *Am J Epidemiol*
643 1990;132:373–5.
- 644 [44] Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-
645 differential misclassification effects: Expectations vs observations. *Int J Epidemiol*
646 2005;34:680–7. doi:10.1093/ije/dyi060.
- 647 [45] Loken E, Gelman A. Measurement error and the replication crisis. *Science* (80-)
648 2017;355:584–5. doi:10.1126/science.aal3618.
- 649 [46] Rutjes A, Reitsma J, Coomarasamy A, Khan K, Bossuyt P. Evaluation of diagnostic
650 tests when there is no gold standard- a review of methods. *Health Technol Assess*
651 (Rockv) 2007;11:1–4. doi:06/90/23 [pii].
- 652 [47] Collins J, Huynh M. Estimation of diagnostic test accuracy without full verification: A

- 653 review of latent class methods. *Stat Med* 2014;33:4141–69. doi:10.1002/sim.6218.
- 654 [48] van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KGM, de Groot JAH. Latent
655 Class Models in Diagnostic Studies When There is No Reference Standard--A
656 Systematic Review. *Am J Epidemiol* 2014;179:423–31. doi:10.1093/aje/kwt286.
- 657
- 658

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

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

