Impact of measurement error in the study of sexually transmitted infections

L Myer, C Morroni, B G Link

Measurement is a fundamental part of all scientific research, and the introduction of errors of different sorts is an inevitable part of the measurement process in epidemiological and clinical research. Despite the ubiquity of measurement error in research, the substantial impacts which measurement error can have on data and subsequent study inferences are frequently overlooked. This review introduces the basic concepts of measurement error that are most relevant to the study of sexually transmitted infections, and demonstrates the impacts of several of the most common forms of measurement error on study results. A self assessment test and MCQs follow this paper.

Reliability and validity

The terms validity and reliability are used in a range of different ways throughout clinical and epidemiological research. In the context of measurement error, these terms refer to the type of error which may be present in a given measure. When we compare a measured value with its true value (or more often, an accepted “gold standard” measure), we assess the validity, or accuracy, of that measure. For variables that can take only two values, such as positive versus negative test results, sensitivity (the proportion of true positives correctly classified as positive by a test) and specificity (the proportion of true negatives correctly classified by a test) are commonly used measures of validity.

In situations where true values of a given variable are not known (or when a widely accepted “gold standard” does not exist), we often compare repetitions of the same measurement to assess the reliability, or precision, of the measure. In these instances reliability is thought of as an indirect assessment of validity, with the reliability of a given measure representing its maximum possible validity. An ideal measure is both valid (accurate) and reliable (precise); while a measure with a high reliability may have a low validity, a measure with a low reliability cannot have a high validity (that is, a single measure which is imprecise cannot be accurate). Various correlations coefficients (for continuous variables) and Cohen’s kappa (for categorical variables) are commonly used measures of reliability.

Types of measurement error

While validity and reliability refer to how a given measure relates to the truth, it is also important to understand how measurement errors may follow a pattern within a study. We use systematic error to refer to any kind of measurement error that leads to systematic (that is, non-random) differences between the observed measurement and its true value. Within this broad definition, systematic error is often used to refer specifically to situations in which the presence, or degree, of measurement error depends on the values of other variables involved in the study. In their most common form, systematic errors occur when measurement errors vary between the groups under comparison in a study (for instance, when the degree of error varies between cases and controls in a case-control study). In this case, systematic errors in measurement give rise to different forms of information bias in epidemiological studies.

If measurement error does not occur systematically within a study, we think of it as random...
error. Random error in the measurement of a variable means that the mistakes are unpredictable, and occur independent of the values of other variables in the study. These errors, which can be thought of as arising because of chance (even if the source of the error is known), apply equally in presence and degree to all groups within a study. One example of this is a laboratory technician’s difficulty in reading a rapid plasma reagin (RPR) test result for syphilis, where distinguishing a weak positive result from a negative result can be difficult and is prone to human error. If the technician is blinded to any information about the research subjects involved, any mistakes are likely to occur independently of other variables involved in the study, and would represent a source of random error.

**Variable form**

In some instances the impact of a particular type of measurement error on the results of a study will depend on the form of the variable involved. A basic distinction should be made between continuous variables (whose value can take any value along a scale, such as age or CD4 cell count) and categorical variables (whose value can only take a limited number of values on a scale, such as infection status). There are further distinctions among categorical variables that are dichotomous or binary (with only two possible values) or polytomous (involving three or more values).

Here we will focus primarily on dichotomous variables, as these are the most commonly encountered in clinical and epidemiological research, and the basic principles of measurement error can be demonstrated simply using concepts of the sensitivity and specificity of a measure; however, it is important to note that the impact of different forms of measurement error varies in some instances according to the variable form. The terms systematic error and random error are often used to refer to continuous variables in particular. For dichotomous variables, the concept of measurement error is often referred to as misclassification; in this framework, systematic error is often thought of as differential misclassification and random error as non-differential misclassification.

**Role of variable in analysis**

The impact of measurement error on study results will depend heavily on what part the variable in question plays in study analysis. A preliminary distinction is made here between independent variables (“risk factors” or “exposures” of interest), dependent variables (“outcomes” or the disease in question), and “confounders” (covariates that are correlated with the exposure of interest and also causally associated with the outcome under study). Measurement error in other variable types (such as effect modifiers or mediators) will not be dealt with here.

**COMMON FORMS OF MEASUREMENT ERROR AND THEIR IMPACT**

Although measurement error can enter into studies and impact their results in a number of ways, the most common forms of measurement error fall into three general categories.

**Non-differential (random) errors in exposure and/or outcome variables**

Non-differential or random errors in an exposure and/or an outcome variable typically cause the categories under comparison to become more similar. This mixing or homogenisation of effects leads to an attenuation or weakening of the observed association between the exposure and the outcome. This is reflected in the observed measure of association (for example, a relative risk or odds ratio) becoming smaller than the true association as the two groups being compared become more similar.

For example, in a hypothetical cohort study of bacterial vaginosis (BV) in pregnant women and risk of low birthweight deliveries, the assessment of BV during pregnancy may be based on the Amsel (clinical) criteria. These criteria may be used as a proxy for the microbiological definition of BV (via the Nugent criteria applied to Gram stains of vaginal fluid), albeit with some degree of error. If the sensitivity and specificity of the clinical criteria for BV are 70% and 95%, respectively, when compared to the microbiological criteria, it is possible to estimate the effect of the measurement error on the resulting relative risk. As example 1 shows (see appendix), the impact of imperfect measures may be substantial, in this case, reducing an appreciable true association to a smaller and statistically insignificant observed association.

Although this example involves random measurement error in an independent variable only, a similar attenuation of the measure of association will arise when this type of error occurs in the measurement of a dependent variable. The degree of attenuation of the observed measure of association will be compounded if there is random error in both independent and dependent variables. In the extreme case, two variables that are measured with complete random error (that is, whose values are determined via some random process) should not be associated at all.

As this example shows, these errors will usually skew study findings to show no association (or a smaller association) between two variables when one is truly present. As a result, non-differential or random error in the measurement of exposure or outcome variables is an important possibility for consideration when a study reports no association between two variables, and this form of random error will need special consideration when the exposure and/or outcome of interest may be difficult to measure precisely. The impact of random errors in polytomous exposure or outcome variables can be more complex, depending on whether some or all of the levels of the variable are involved.

**Differential (systematic) errors in exposure and/or outcome variables**

Errors in the measurement of an exposure and/or an outcome variable can occur differently for various groups within a study. Such systematic or differential errors often result from the types of information bias that most researchers are familiar with (recall bias, diagnostic bias, etc.). Although researchers are most concerned with differential errors which may artificially inflate study results (for example, a recall bias that skews the odds ratio in a case-control study away from the null value), it is important to note that systematic measurement errors can act to either attenuate or inflate the measure of association between an exposure and a disease; the precise effect of these errors will depend on how the errors operate in a given context. For more on various biases and their potential impacts, readers are referred to introductory epidemiology or clinical trials textbooks with sections focusing specifically on bias.

**Errors in the measurement of confounders**

Most epidemiological and clinical research devotes considerable effort to measure exposure and disease variables as accurately as possible. With this focus, less attention is given typically to the rigorous measurement of covariates which may act as confounders of the exposure-disease association of interest. These variables are commonly included in multivariate statistical analyses (for example, logistic regression), with researchers reporting measures of the exposure-disease association as being adjusted for particular covariates. However, when confounding variables are measured with error, the resulting statistical adjustments will be imperfect,
and will not remove completely the true confounding effect.
The result will be the appearance of an “adjusted” measure of
association which is still truly confounded; in the extreme
case, an exposure-disease association will appear to exist
after adjustment when none is in fact present.4

For example, many investigations into the aetiology of HIV
and other sexually transmitted infections involve measuring
and adjusting for high risk sexual behaviours as potential
confounding variables. Any measurement of sexual beha-
vour is likely to be subject to at least some measurement
error, if only because of reporting errors by participants.2
In a possible cohort study of hormonal contraception use
and risk of HIV infection, high risk sexual behaviours may be
associated with contraception as well as HIV infection,
making sexual behaviour measures potential confounders
be adjusted for.9 10 In a hypothetical “true” scenario
(example 2, appendix), it is plausible that an appreciable
unadjusted relative risk (RR) of over 2.0 could be completely
confounded by sexual behaviours (scenario 1, appendix). Yet
even relatively small mismeasurement of risk behaviours will
lead to an inability to adequately adjust for the true
confounding effects, creating in turn the false appearance
of an “adjusted” association. Note that the degree of
mismeasurement in this example (with a sensitivity of 75%
and specificity of 80%) is relatively minor; although there are
few studies on the validity of self reported sexual behaviours,
the amount of error may be far greater—further reducing the
ability to adjust for true confounding effects.11 12

This phenomenon—sometimes referred to as “residual
confounding”—is probably far more common than is widely
recognised.13 It is likely to be especially problematic when
the variable being measured is difficult to quantify, such as
measurements of socioeconomic status, or when the measure
only captures one facet of the variable of interest—for
instance, when indirect serological tests are used to measure
the presence of some bacterial infections. The possibility of
residual confounding requires consideration in any situation
where an exposure-disease association persists after statis-
tical adjustment for a known confounder. Particular care
should be given to the possible role of random errors in the
measurement of exposure and/or disease (which may help
to explain the lack of an observed association between two
variables), as well as in the measurement of confounding
variables (which may help to explain the persistence of an
observed exposure-disease association despite apparent
“adjustment”).

CONCLUSION
The different forms of measurement error described here are
a commonly overlooked aspect of research involving sexually
transmitted infections. However, as epidemiological and
clinical sciences attempt to detect more subtle associations,
increasing attention is being paid to the part that measure-
ment error may play in distorting study results.17 18 In order to
counter the impact that such errors may have, researchers
need to understand the different types of measurement error,
be able to recognise sources of measurement error in a given
study, and reflect critically on the potential impact of
measurement error on research findings. Ultimately, the
explicit recognition of measurement error in research,
accompanying by an evaluation of the robustness of data in
the face of such measurement error, leads to stronger, more
interpretable, and more defensible scientific results.

ACCOUNTING FOR THE EFFECTS OF MEASUREMENT
ERROR
Researchers can attempt to minimise the potential impact of
measurement error on their study results at different stages
of the research process. When designing a study, it is critical
for scientists to consider the validity or reliability of the
measures they plan to employ. If the measures being used are
not accepted gold standards, it is important to understand
how the measure may relate to the gold standard. If this
information is not known, then a small validation substudy
may be warranted; if no gold standards exist for the
variable(s) being measured then researchers should seek to
evaluate the reliability of the measure as a proxy for its
validity.

In data analysis, it may be possible to correct for simple
forms of measurement error if the validity of the measures
involved has been established. There are a range of
established statistical methods for these adjustments, and
in their simplest form these can be quite user friendly.15
However they often require a range of assumptions that may
be untestable, and the calculations involved quickly become
complex in multivariate analyses, particularly when the
possibility of correlations between errors in different vari-
ables may be impossible to predict.16

More generally, it is possible to estimate the effect that
different degrees of measurement error (whether systematic
or random) may have had on the observed data from a given
study. These estimations, sometimes referred to as sensitivity
analyses, parallel the examples used here. Such simple simu-
lations can be very useful for assessing the robustness of
study findings in general, or quantifying the plausible impact
that specific errors could have in explaining the observed
results. Given the relatively accessibility of basic sensitivity
analyses, it may be useful to include such analyses in a
research report or journal publication as part of evaluating
the strengths and weaknesses of a study’s findings.

Finally, in reporting study results researchers should
provide a straightforward assessment of the different forms
of error that may have been present in their study, and how
these may affect study inferences. While the more widely
known systematic errors that bias study results are usually
addressed by most investigators, special attention should also
be given to the possible role of random errors in the
measurement of exposure and/or disease (which may help
to explain the lack of an observed association between two
variables), as well as in the measurement of confounding
variables (which may help to explain the persistence of an
observed exposure-disease association despite apparent
“adjustment”).

Authors’ affiliations
L Myer, School of Public Health and Family Medicine, University of Cape
Town, Cape Town, South Africa
C Marroni, Women’s Health Research Unit, School of Public Health and
Family Medicine, University of Cape Town, Cape Town, South Africa
B G Link, Department of Epidemiology, Mailman School of Public
Health, Columbia University and New York State Psychiatric Institute,
New York, USA

REFERENCES
1 Greenland S. The effect of misclassification in the presence of covariates.
2 Barron B. The effects of misclassification on the estimation of the relative risk.
3 Schwebke JR, Hillier SL, Sobel JD, et al. Validity of the vaginal gram
4 Macleir M, Schneeveis S. Causation of bias: the epidemleiology
5 Rahman K, Greenland S. Modern epidemiology. Philadelphia: Lippincott
Williams & Wilkins, 1998.
6 Phillips AN, Davey Smith G. Bias in relative odds estimation owing to

www.stijournal.com
Multiple choice questions (see p 328 for answers)

1. Information bias is one form of measurement error.  
(a) True  
(b) False.

2. The sensitivity of a measure is an assessment of its:  
(a) Reliability  
(b) Positive predictive value  
(c) Validity  
(d) Precision.

3. The key difference between systematic error and random error is that:  
(a) The cause of systematic error is known, while the origins of random error are unknown  
(b) Systematic errors are patterned in data according to the values of other variables, while random errors are not  
(c) Random error does not affect the results of a study, but systematic error does  
(d) Systematic error can also be thought of as non-differential misclassification, while random error can be thought of as differential misclassification.

4. Random errors in the measurement of a binary exposure variable in a study will most commonly lead to:  
(a) No change in the study results  
(b) An increase in the observed measure of association, so that the exposure-disease association appears stronger than it truly is  
(c) A decrease in the observed measure of association, so that the exposure-disease association appears weaker than it truly is  
(d) The impact of this form of measurement error is highly unpredictable and can not be generalised.

5. Random errors in the measurement of a binary outcome variable in a study will most commonly lead to:  
(a) No change in the study results  
(b) An increase in the observed measure of association, so that the exposure-disease association appears stronger than it truly is  
(c) An increase in the unadjusted (crude) association between exposure and disease  
(d) A reduced ability to control in statistical analysis for the confounding variable.

6. Understanding how measurement error may impact on study results is important because:  
(a) It is possible to minimise measurement error in designing a study by understanding the validity or reliability of the measures to be used  
(b) It is possible to assess the potential impact of measurement error in data analysis through simple sensitivity analyses  
(c) Understanding measurement error and its possible effects is helpful in critically evaluating published research  
(d) All of the above.

7. The exact impacts of different forms of measurement error can vary slightly between binary, polytomous and continuous variables.  
(a) True  
(b) False.

8. Thinking about the potential impacts of measurement error is most important for which kind of variables?  
(a) Behavioural variables  
(b) Laboratory values  
(c) Clinical measures  
(d) All of the above.

9. Correlation coefficients and Cohen’s kappa are used to assess the:  
(a) Validity of a measure  
(b) Reliability of a measure  
(c) Accuracy of a measure  
(d) Specificity of a measure.
Impact of measurement error in the study of sexually transmitted infections

L Myer, C Morroni and B G Link

Sex Transm Infect 2004 80: 318-321
doi: 10.1136/sti.2003.006536

Updated information and services can be found at:
http://sti.bmj.com/content/80/4/318

These include:

References
This article cites 15 articles, 4 of which you can access for free at:
http://sti.bmj.com/content/80/4/318#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/