PEER-REVIEWED

Why you should read this article:

- Measurement error must always be considered when planning a research project and interpreting its results
- Little attention has been given in nursing research training to misclassification and measurement errors
- Ignoring measurement error will result in flawed interpretation and application of study results to clinical practice

Misclassification and measurement error – planning a study and interpreting results

Steven Alfred Frost, Evan Alexandrou

Citation
Frost SA, Alexandrou E (2021) Misclassification and measurement error – planning a study and interpreting results. Nurse Researcher. doi: 10.7748/nr.2021.e1765

Peer review
This article has been subject to external double-blind peer review and has been checked for plagiarism using automated software

Correspondence
s.frost@westernsydney.edu.au

Conflict of interest
None declared

Accepted
14 August 2020

Published online
January 2021

Abstract

Background Measurement error must always be considered when planning a research project and interpreting its results. The accuracy of some data collected during a study can often be confidently assured, but more than one measurement or observer is needed to assess exposure and outcomes status in cases where clinical measurement is prone to measurement error. Little attention is paid in nursing research to misclassification and measurement error. Bias is often discussed in nursing research education, but not its potential consequences or measures that can be taken to improve the study's quality.

Aim To present examples of random measurement error – misclassification of a binary outcome – in a continuous exposure and outcomes variable, to address this gap in nurses' research training.

Discussion The article discusses the relationship between exposure and outcome in the absence and presence of measurement error using risk (relative risk) and association using correlation. It provides methods to estimate the true value of these measures of risk and association, when only given the clinical measurements with errors.

Conclusion If the assumption of random error holds, attenuation of risk or association towards the null will occur.

Implications for practice Understanding the effect of measurement error including misclassification will enable researchers to interpret the results of their studies, and to take into consideration this potential error when planning and conducting a study.

Author details

Steven Alfred Frost, deputy director, South Western Sydney Local Health District, Centre for Applied Nursing Research, Liverpool BC, New South Wales, Australia

Evan Alexandrou, associate professor, School of Nursing, Western Sydney University, Penrith, New South Wales, Australia

Keywords

data collection, data analysis, quantitative research, research, study design

Introduction

Measurement error must always be considered when planning a research project and interpreting its results. The accuracy of some data collected during a study can often be confidently assured, but more than one measurement or observer is needed to assess exposure and outcomes status in the case of a clinical measurement that is prone to measurement error.

Little attention is paid in nursing research to misclassification and measurement error. Bias is often discussed in nursing research education, but not its potential consequences or measures that can be taken to improve the study's quality.

This article will provide a brief background and use clinical examples to show the effects of misclassification and measurement errors in the results obtained from a study.

Misclassification of exposure status

To demonstrate misclassification, we will present a fictional example that uses delirium as an exposure and a fall as the outcome, both of which are binary variables: yes or no. All acute hospitals across our local health district use the 'Confusion Assessment Method' (CAM) (Inouye et al 1990) to identify acute episodes of delirium in patients who appear to be disorientated or confused or whose behaviour or level of consciousness changes.

The CAM was developed using the Diagnostic and Statistical Manual of Mental Disorder (DSM III) (American Psychiatric Association 1980) as the gold standard, so we have good information about how well it identifies individuals with confirmed delirium ('sensitivity') and what proportion of people without delirium it will correctly classify as delirium-free ('specificity'). Shi et al's (2013) large meta-analysis estimated that it has a sensitivity of 0.80 and a specificity of 0.99, so most errors are false negatives, while few are false positives.

Table 1 shows the results of a hypothetical study investigating the relationship between delirium and the risk of fall. The study involved 300 participants, 20% (*n*=60) of whom were observed to have a fall and 33% (*n*=100) of whom were classified using CAM as having had an episode of delirium. The observed rate of falls among those with delirium was 40% (*n*=40) and was 10% (*n*=20) among those without delirium (*n*=200). That means the relative risk was 4.0 (95% CI: 2.48, 6.48).

As we have a gold standard measurement, we can estimate the true rate of delirium (P) from the rate observed with CAM (p) by using Kelsey et al's (1986) formula, P = (p + specificity - 1) / (sensitivity + specificity - 1). In cases where such a gold standard measurement is unavailable, repeated tests of reliability and agreement – usually with at least two blinded independent raters to assess interrater reliability – can be used to determine an estimated true rate using a Kappa statistic (Cohen 1960).

In the case of this study, P = (p – 0.01)/(0.79), so we can see that among fallers, the estimated true rate of delirium is 83% (0.67-0.1/0.79) and among non-fallers is 30% (0.25-0.01/0.79); overall, it is 41% (0.33-0.01/0.79), with 123 estimated to have delirium. The estimated true rate of falls among those with delirium is therefore 50/123 (41%) and 10/177 (5.6%) among those without delirium (Table 2). That means the estimated true relative risk is 7.20 (95% CI: 3.80-13.63).

Table 1. Observed rates of delirium among faller and non-fallers

Delirium	Fallers		Non-fallers		Total	
	n	%	*n*	%	*n*	%
Yes	40	67	60	25	100	33
No	20	33	180	75	200	77
Total	60	20	240	80	300	100

Table 2. Estimated true rates of delirium among faller and non-fallers

Delirium	Fallers		Non-fallers		Total	
	n	%	*n*	%	*n*	%
Yes	50	83	73	30	123	41
No	10	17	167	70	177	59
Total	60	20	240	80	300	100

Key points

- Measurement error and misclassification is almost unavoidable during any study project
- Non-differential or random error will underestimate observed effects and bias results toward the null
- The effect of systematic error or non-random error can bias observed effects away or toward the null

Measurement error of a continuous variable

Most clinical measurements will demonstrate some within-individual variation (error), typically estimated using repeat measurements over a short period of time to avoid any underlying biological variation. Measurement error usually assumes a normally distributed clinical variable and an associated similar random error, with a mean of zero and a standard deviation or variance obtained from repeat measurements within-an-individual.

Example 1

The following example uses a subset of the Framingham Heart study data (Dawber et al 1951, National Heart, Lung, and Blood Institute 2010) to show the association between total serum cholesterol (TC) (mg/dL) and systolic blood pressure (SBP) (mmHg). Figure 1 shows the correlation between the observed TC and SBP without errors – the correlation is 0.20, and the associated ordinary least squares (OLS) slope is 0.10.

Figure 1. Association between total cholesterol and systolic blood pressure measurements without errors

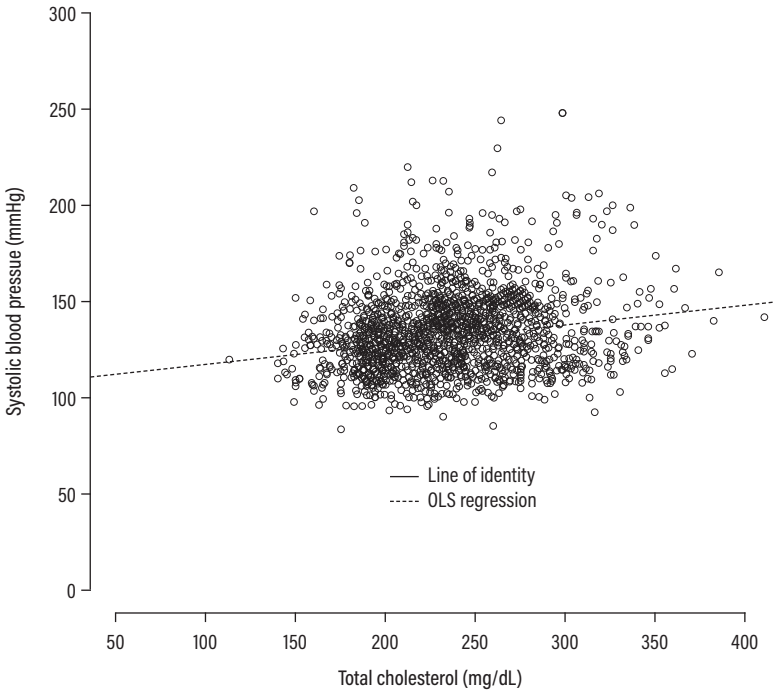

Figure 2 shows the correlation when random error is added to TC with mean 0mg/dL (SD 25mg/dL) and to SBP with mean 0mmHg (SD 15mmHg) – the correlation becomes 0.13 and the OLS slope becomes 0.07.

The true correlation can be estimated based on the measurements without error using the following formula (Kelsey et al 1986):

$$rho_{true} = rho_{observed}/(rho_{xX} \text{ X } rho_{yY})$$

where, rho_{true} is the estimated Pearson correlation coefficient between the measurements (X, Y) without error, $rho_{observed}$ the Pearson correlation coefficient between (x, y) with error, and rho_{xX} and rh_{yY} are the correlations between the measurements without and with errors. In this case, x=TC and y=SBP, $rho_{observed}$ is 0.13, rho_{xX} is 0.88 and rho_{yY} 0.83, the latter two values being obtained by estimating the correlation between the respective measurements with and without error. Therefore, the estimated true correlation rho_{true} is 0.18.

Example 2

To further show the effect of random measurement error, we will examine the relationship between TC, while adjusting for age and sex, and the risk of systolic hypertension (SH) – that is, of having an SBP greater than 150mmHg. For simplicity, we present the case where the error is random.

Table 3 shows the risk of SH based on age, being male and each 18mg/dL increase in TC, for TC and SBP without error; Table 4 shows the same information with error. Without error, after adjusting for age and being male using a Poisson regression (Breslow and Day 1980), an 18mg/dL increase in TC increases the relative risk of SH by 6% (95% CI: 0%, 11%); when random error is added to SBP and TC, the increase is only 1% (95% CI: -3%, +5%).

Discussion

We have given examples of the effect of measurement error and misclassification of the results of assessing the relationship

between an exposure and an outcome of interest. If the error is random, the movement of estimates of effect towards the null is obvious.

This effect of error not just in the exposure factor but also in the outcome of interest should always be considered when interpreting the results of a study. Importantly, in the context of random measurement error and misclassification, the effect of an exposure will be underestimated – this has implications for the sample size needed and the approach used to collect data.

The potential effect of measurement error on the estimation of the relationship between an exposure and outcome has been explored to a great extent in epidemiological studies of risk, particularly nutritional epidemiology (Greenland and Kleinbaum 1983, Willett 1989, Rosner 1996). In many cases, measurement error can be quantified by repeated within-individual measurements and will give a good guide of the precision of a given measurement (Irwig et al 1991).

Our examples highlight the effect of random error and its common consequences on measures of effect. Importantly, careful planning has enabled researchers to address measurement error during pilot phases, by developing good estimates of the precision of exposure measures that often involve very detailed data collection with a small subset of the study population (Bingham et al 1997, Day et al 1999), allowing the results of the final study's outcomes to be corrected for measurement error (Riboli et al 2002).

Many clinical measurements have been extensively explored, using repeated, within-individual measurements, to identify the magnitude of precision, which is presented in many cases as the coefficient of variance: the ratio of the standard deviation of within-individual repeated measures and the population mean. The estimated precision of bone mineral density (BMD) is less than 2%; it is of the order of 10% for serum cholesterol (Irwig et al 1991). The attenuation of the risk of fracture and BMD would therefore be small when compared to that of an increase in serum cholesterol level and risk of stroke.

Importantly, researchers must acknowledge when it is known there is

Figure 2. Association between total cholesterol and systolic blood pressure measurements with errors

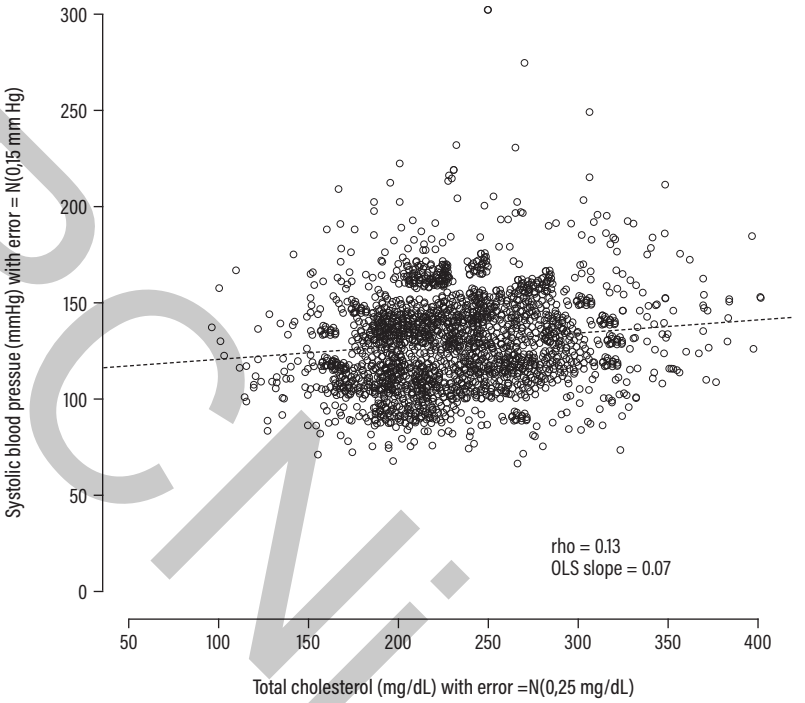

Table 3. Estimates of relative risk of SH without error

	Crude	CI 95%	*P*-value	Adjusted	CI 95%	*P*-value
Age (10+ years)	2.14	1.84, 2.48	<0.001	2.05	1.76, 2.40	<0.001
Male	0.73	0.57, 0.94	0.015	0.79	0.61, 1.02	0.069
18mg/dL TC increase	1.13	1.08, 1.18	<0.001	1.06	1.00, 1.11	0.044

Table 4. Estimates of relative risk of SH with error

	Crude	CI 95%	*P*-value	Adjusted	CI 95%	*P*-value
Age (10+ years)	1.85	1.62, 2.10	<0.001	1.83	1.60, 2.09	<0.001
Male	0.72	0.58, 0.90	0.004	0.75	0.60, 0.93	0.010
18mg/dL TC increase	1.05	1.01, 1.09	0.009	1.01	0.97, 1.05	0.773

error in a given clinical measurement that this could lead to an underestimation of risk. Measurement error can also increase the risk of Type II error, so a greater sample size may be needed for a given trial.

When error is not random, it is difficult to estimate the direction of bias. This is why it is imperative when planning a study to consider measurement error and ask: is the error random or nondifferential between exposure and outcomes groups? Ways to reduce the potential effect of measurement error include blinding of exposure status when assessing outcome and blinding of outcome status when assessing exposure in a case-control study. Randomisation will assign error evenly between intervention and control groups, and is the most obvious approach to reduce bias. However, measurement error will need to be carefully mitigated in some cases, using repeated measurements or multiple raters, or direct estimation using a small (random) subset of participants. Once the level of within-individual error has been estimated, this can be used to correct measures of effect (Willett 1989).

Conclusion

We have presented some examples of the effects of measurement error and the estimates of risk or association. Importantly, if the assumption of random error holds, the attenuation of risk or association will move toward the null. Understanding the effect of measurement error including misclassification will enable researchers to interpret the results of a study and consider this potential error when planning and conducting research.

References

American Psychiatric Association (1980) Diagnostic and Statistical Manual of Mental Disorders. Third edition. APA, Philadelphia PA.

Bingham SA, Gill C, Welch A et al (1997) Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. International Journal of Epidemiology, 26, Suppl 1, S137-S151. doi: 10.1093/ije/26.suppl_1.s137.

Breslow NE, Day NE (1980) Statistical Methods in Cancer Research. International Agency for Research on Cancer, Lyon, France.

Cohen J (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 20, 1, 37-46. doi: 10.1177/001316446002000104.

Day N, Oakes S, Luben R et al (1999) EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. British Journal of Cancer. 80, Suppl 1, 95-103.

Dawber TR, Meadors GF, Moore FE Jr (1951) Epidemiological approaches to heart disease: the Framingham Study. American Journal of Public Health and the Nation's Health. 41, 3, 279-281. doi: 10.2105/ajph.41.3.279.

Greenland S, Kleinbaum DG (1983) Correcting for misclassification in two-way tables and matched-pair studies. International Journal of Epidemiology, 12, 93-97. doi: 10.1093/ije/12.1.93.

Inouye SK, van Dyck CH, Alessi CA et al (1990) Clarifying confusion: the confusion assessment method. A new method for detection of delirium. Annals of Internal Medicine. 113, 12, 941-948. doi: 10.7326/0003-4819-113-12-941.

Irwig L, Glasziou P, Wilson A et al (1991) Estimating an individual's true cholesterol level and response to intervention. JAMA. 266, 12, 1678-1685. doi: 10.1001/jama.1991.03470120080037

Kelsey JL, Thompson WD, Evans AS (1986) Methods in Observational Epidemiology. Oxford University Press, New York NY.

National Heart, Lung, and Blood Institute (2010) Teaching Datasets. https://biolincc.nhlbi.nih.gov/teaching/

Riboli E, Hunt KJ, Slimani N et al (2002) European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutrition. 5, 6B, 1113-1124. doi: 10.1079/PHN2002394.

Rosner BA (1996) Measurement error models for ordinal exposure variables measured with error. Statistics in Medicine. 15, 3, 293-303. doi: 10.1002/(SICI)1097-0258(19960215)15:3<293::AID-SIM166>3.0.CO;2-E.

Shi Q, Warren L, Saposnik G et al (2013) Confusion assessment method: a systematic review and meta-analysis of diagnostic accuracy. Neuropsychiatric Disease and Treatment. 9, 1359-1370. doi: 10.2147/NDT.S49520.

Willett W (1989) An overview of issues related to the correction of non-differential exposure measurement error in epidemiologic studies. Statistics in Medicine. 8, 9, 1031-1040. doi: 10.1002/sim.4780080903.