Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: informahealthcare.com/journals/imte20

Better data ≫ Bigger data

Kulamakan Kulasegaram & Elise Paradis

To cite this article: Kulamakan Kulasegaram & Elise Paradis (2014) Better data ≫ Bigger data, Medical Teacher, 36:11, 1008-1009, DOI: 10.3109/0142159X.2014.917761

To link to this article: https://doi.org/10.3109/0142159X.2014.917761

Published online: 29 Jul 2014.

with reliability and actually is one of the common mistakes in reliability analysis (Rothman et al. 2008). Reliability (repeatability or reproducibility) is often assessed by different statistical tests such as Pearson r, least square and paired t. 'Mistakes in reliability analysis are common' (Lawrence & Kuei 1989; Rothman et al. 2008).

For quantitative variables the Intra Class Correlation Coefficient (ICC) should be used. For qualitative variables the weighted kappa, which should be used with caution because kappa has its own limitation too (Lawrence & Kuei 1989; Rothman et al. 2008). It is crucial to know that there is no value of kappa that can be regarded universally as an indication good agreement. An important weakness of *k* value to assess agreement of a qualitative variable is that it depends upon the prevalence in each category. This means that it is be possible to have a different kappa value based on the same percentage of both concordant and discordant cells.

The authors point out in their conclusion, "peer raters" of the same level of training can provide accurate ratings of complex clinical tasks and can serve as an important resource in assessing student performance in an OSCE, but have not investigated the concordance of the pass/fail decisions with respect to individual candidates

Reliability (precision) and validity (accuracy) are two completely different and important methodological issues in all fields of researches. To assess the accuracy (validity) the following tests are used:-
sensitivity (the percentage with the disease who test positive, True Positives / (True Positives + False Negative)),
specificity (the percentage of healthy who test negative, True Negatives / (True Negatives + False Positive))
positive predictive value (PPV), (percentage of positive tests who actually are diseased, True Positives / (True Positives + False Positive)),
negative predictive value (NPV) (the percentage of negative tests who are healthy, True Negatives / (True Negatives + False Negative)),
likelihood ratio positive and likelihood ratio negative as well as diagnostic accuracy [(both true positive and true negative results / total)× 100]
odds ratio (true results / false results) preferably more than 50.

These are the tests to evaluate the validity (accuracy) of a test compared to a gold standard (Rothman et al. 2008).

Therefore, the authors' conclusion is due to the confusion of reliability (precision) with validity (accuracy) and is, therefore, misleading.

Siamak Sabour, MD, MSc, DSc, PhD, Postdoc, Department of Clinical Epidemiology, Faculty of Health, Shahid Beheshti University of Medical Sciences, Tehran, I.R. Iran. Tel: +98 21 22421814. E-mail: s.sabour@sbmu.ac.ir.

Declaration of interest: The author reports no conflicts of interest.

References

Basehore PM, Pomerantz SC, Gentile M. 2014. Reliability and benefits of medical student peers in rating complex clinical skills. Med Teach 2014;36(5):409–414.

Lawrence I, Kuei L. 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45:255–268.

Rothman JK, Greenland S, Timothy LL. 2008. Modern epidemiology, 3rd ed. Baltimore, USA: Lippincott Williams & Wilkins.

Re: Reliability and benefits of medical student peers in rating complex clinical skills: Response to common mistake

Dear Sir

We want to take this opportunity to respond to the concerns raised about the reliability analysis conducted in the study. Dr Sabour has pointed out the appropriate use of intraclass correlation coefficient (ICC) as a preferred analysis to assess reliability in quantitative variables and has criticized our use of Pearson correlation coefficient.

As Dr Sabour is likely aware, the G-coefficient in generalizability analysis and ICC are both based in classical test theory and are closely related. While ICC analysis examines a single facet, generalizability analysis provides the opportunity to look at multiple facets of measurement error in a single design (Shrout & Fleiss 1979; Barch & Mathalon 2011). In our analysis, while the correlation coefficients were used to establish the relationship between peer and faculty ratings, the generalizability analysis provided the reliability measure.

We appreciate his interest in our research and the opportunity to clarify the analysis conducted.

Pamela M. Basehore and Sherry C. Pomerantz, Rowan University School of Osteopathic Medicine, Academic Affairs, 1 Medical Center Drive, Stratford, New Jersey 08084, USA. E-mail: basehore@rowan.edu

Declaration of interest: The authors report no conflicts of interest.

References

Barch DM, Mathalon DH. 2011. Using brain imaging measures in studies of precognitive pharmacological agents in schizophrenia: Psychometric and quality assurance considerations (Supplement). Biol Psychiatry 70(1):13–18.

Basehore PM, Pomerantz SC, Gentile M. 2014. Reliability and benefits of medical student peers in rating complex clinical skills. Med Teach 36(5):409–414.

Shrout PE, Fleiss JL. 1979. Intraclass correlations: Uses in assessing rater reliability. Psych Bull 86(2):420–428.

Better data ≫ Bigger data

Dear Sir

We read Ellaway et al.'s (2014) article on Big Data in health professions education with great interest. We share the authors' excitement and thank them for starting the conversation in our field. Here we stress two key Big Data concerns: while analytics have undeniable benefits for hypothesis generation, we can't eschew broader questions of scientific design and analysis.

First, Big Data is not objective data. Just as with small, purposeful datasets, large datasets are defined by the assumptions, questions, tools, and interpretations that underpin them. Our understanding of health professions education may regress if we ignore issues of design, construct selection and validation of measurements. Large or small, purposefully collected datasets wrestle with these issues upfront; datasets of convenience rarely do.

Second, not all data analysis – no matter how large the dataset – constitutes science. Exploration of the signals (and noise) in large datasets without adequate conceptual frameworks can be misleading if not dangerous. Secondary data analysis is a useful but inherently limited scientific tool as it cannot robustly infer causation. It is only when data collection and analysis are informed by theory that robust results are possible.

The scientific method was developed to navigate the complex challenges of making meaning from data. In this endeavor, better data will always trump bigger data. Without proper design and analytic rigor, Big Data could easily make us aggrandize spurious results and lead us astray.

Others fields have navigated these challenges and used theory to guide Big Data. For example, Shwed and Bearman (2010) used Latour's 'Black Box' theory to model scientific consensus formation. They analyzed citation networks from about 30,000 publications and 124,000 citations to shed light on controversies such as the carcinogenicity of tobacco and the autism/MMR vaccine connection. In medical education, Asch and colleagues (2009) tracked maternal complication rates for 4000 obstetricians who collectively performed 4.9 million deliveries over 15 years. The authors showed the effects of training program, experience, and individual ability on clinical performance, thereby testing and confirming theories developed by experimental studies.

These studies suggest that we as a community of scholars can use Big Data to *serve* research, rather than have Big Data *dictate* it. Meaningful knowledge comes only from scientifically informed design and analysis. Ultimately, it is *not* about the size of the dataset.

Kulamakan Kulasegaram, Elise Paradis, The Wilson Centre, University of Toronto, Toronto, Ontario, Canada. E-mail: mahan.kulasegaram@utoronto.ca

References

Asch DA, Nicholson S, Sindhu S, Herrin J, Epstein AJ. 2009. Evaluating obstetrical residency programs using patient outcomes. JAMA 302(12):1277–1283.

Ellaway RH, Pusic MV, Galbraith RM, Cameron T. 2014. Developing the role of big data and analytics in health professional education. Med Teach 36(3):216–222.

Shwed U, Bearman PS. 2010. The temporal structure of scientific consensus formation. Am Sociol Rev 75(6):817–840.

Re: 'Better data ≫ Bigger data'

Dear Sir

We thank Drs. Kulasegaram and Paradis for their considered letter, and the addition to the emerging discourse around Big Data in health professions education. We would like to respond to a few of the specific points they make.

We agree that 'datasets of convenience' should be considered in terms of their objectivity, provenance, and semantic baggage. We had hoped, in preparing the original paper, to provoke a debate on the extent to which data of uncertain provenance or applicability may be used to make decisions that have serious consequences for students, faculty and others in medical education. The expectation that data collected in one context and for one purpose can subsequently be used in and for others should always be tested, both theoretically and empirically. As Big Data begins to be used in health professional education we need to ensure that it is done in a critical and scholarly way. It is not just that the data potentially lacks objectivity and theoretical grounding (a problem for research as a whole); it is also that the practices of Big Data may be found wanting, particularly if they develop in isolation.

We would re-emphasize that, as we stated in the original paper, "traditional and Big Data methods should not be considered as solitudes but rather as different approaches that can be productively combined". We urge scholars to explore how Big Data techniques can be meaningfully added to the academic repertoire so that analysts and researchers can use them along with other tools and methods to suit their needs and resources.

Health professional education research is a wide field with many intersecting research paradigms. While some research questions undoubtedly depend on better data rather than bigger data, others may need the warts and all messiness of "datasets of convenience" to explore and understand the systems that generate them. The indicators of quality for Big Data scholarship therefore need to relate to the purpose of inquiry as well as the resources it uses.

It would have been hard to select better examples of a Big Data approach than those suggested. For us the key point they make is that their questions were answered by using Big Data in scientific and scholarly ways rather than in ways that were distinct from academic practice. We hope that this trend extends to health professional education.

Rachel Ellaway, Human Sciences Division, Northern Ontario School of Medicine, 935 Ramsey Lake Road, Sudbury, Ontario P3E 2C6, Canada. E-mail: rellaway@nosm.ca

Martin Pusic, Division of Educational Quality and Analytics, New York University, New York, USA.

Robert Galbraith, National Board of Medical Examiners, Philadelphia, Pennsylvania, USA.

Terri Cameron, Association of American Medical Colleges, Washington, District of Columbia, USA.