



Re: 'Better data >> Bigger data'

Rachel Ellaway, Martin Pusic, Robert Galbraith & Terri Cameron

To cite this article: Rachel Ellaway, Martin Pusic, Robert Galbraith & Terri Cameron (2014) Re: 'Better data >> Bigger data', Medical Teacher, 36:11, 1009-1009, DOI: [10.3109/0142159X.2014.926003](https://doi.org/10.3109/0142159X.2014.926003)

To link to this article: <https://doi.org/10.3109/0142159X.2014.926003>



Published online: 17 Jun 2014.



Submit your article to this journal [↗](#)



Article views: 462



View related articles [↗](#)



View Crossmark data [↗](#)

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.