
FROM THE EDITORS' DESK

Open and Abundant Data is the Future of Rehabilitation and Research



Duncan R. Babbage, PhD

From the Centre for Person Centred Research, Auckland University of Technology, Auckland, New Zealand.

Abstract

Development of our current research practices has been driven by a number of assumptions and from operating within practical constraints. Technological change is beginning to remove many of these limits, although our research and practice has so far only gradually and partially evolved in response. The U.S. federal government is now mandating open data repositories for research that it funds. Policy changes regarding open data repositories and an increasing abundance of data arising from both research and practice provide the opportunity to revisit some assumptions. With abundant sources of data that may increasingly be collected automatically during rehabilitation, it seems fundamentally flawed that the resolution of the primary quantitative analysis approaches widely understood in our field is so limited by the need to contain the risk of false positives. Identification of more sophisticated approaches to our data, which may well already exist in the statistical literature, is a high priority. Archives of Physical Medicine and Rehabilitation 2014;95:795-8

© 2014 by the American Congress of Rehabilitation Medicine

Open Data

In May 2013, the U.S. federal government mandated that both publications and datasets resulting from all research they fund (through agencies with grant pools of \geq \$100 million) must be made openly accessible in machine-readable formats.¹ Funding agencies have been required to rapidly prepare implementation plans.² This policy reflects wider calls for researchers to provide open online access to their data,^{3,4} including a 2004 declaration signed by 33 countries regarding access to research data from public funding.⁵ It is a practice that has already become standard in some fields (eg, economics) and is a publication requirement of some journals. Many open research data repositories already exist: one website indexes 602 such online repositories.⁶ Vanishingly few studies in rehabilitation have used such services, though given the lead from the U.S. federal government, which may well be followed elsewhere, this is coming.

Abundant Data

It is not only open data repositories that mean researchers will need to become familiar with working with large datasets—there is reason to believe that data will become increasingly abundant.

Twenty-five years ago, conducting a literature review or looking for subsequent citations to an article included laborious manual review of printed indexes,⁷ locating in a library the paper copy of the issue of the journal in question, and spending many hours over a photocopier to duplicate articles of interest. Source articles were hard wrung from our academic archives. Today, electronic databases and full-text articles provide improved access to the latest knowledge in the field.^{8,9} However, high-quality data remains a scarce resource. Recruitment of participants appears to be a constant challenge for nearly all research studies, particularly in rehabilitation,¹⁰ and the burden for both participants and researchers to collect quality data is high. Current technological developments may be the initial tremor that heralds a tsunami of data heading to our shores. The multiple real-time data streams that can be collected by smartphones, wearable devices, and ambient sensors embedded in the environment¹¹ may soon provide us with access to more data on rehabilitation practice than we are equipped to process, let alone interpret. Furthermore, not only may such data be collected for formal research studies, but in the future it could be collected as part of the standard clinical record of every person receiving rehabilitation—ready and waiting for intelligent analysis. Deep, complete, and arguably invasive clinical records would need careful consideration of issues of privacy, consent, and data security. This data would, however, also provide the opportunity for insight into issues previously opaque, such as enabling clearer connections to be drawn between the actual treatment dose that has been received and

No commercial party having a direct financial interest in the results of the research supporting this article has conferred or will confer a benefit on the author or on any organization with which the author is associated.

outcome—through, for example, knowing precisely how much a compensatory device was actually used in the community. In order to draw meaningful conclusions from this abundance of data, our analysis methods will need to evolve alongside our data collection.

New Analysis Approaches

Topography rather than biopsy

Null hypothesis testing has been a major quantitative research tool. Statistical techniques (eg, *t* tests, F tests) have provided answers regarding effective interventions. However, such techniques suffer from major drawbacks, most notably that the entire premise is the ability to be 95% confident about a single specific analysis. This limitation has long been recognized, as has the associated inherent risks in undertaking multiple analyses: with a 5% chance of a spurious result from any one analysis, the investigations conducted with a dataset must, of necessity, be selected in advance and be carefully limited. A reasonable analogy is taking biopsies. In the absence of any other tool to determine the nature of a condition, a biopsy is a necessary but intrusive and potentially risky operation; therefore, it is always important to take as few biopsies as possible. The use of statistical significant testing seems analogous to taking a biopsy of a dataset; it is a penetrating analysis but needs to be used as sparingly as possible to contain the risk of false positives. Beginning to address this, we have moved our quantitative analyses beyond simple statistical significance testing to the reporting of associated effect sizes. We have a range of more sophisticated multivariate analyses, such as analyses using the generalized linear model and structural equation modeling. The complexity of the relationships between the variables that we examine in rehabilitation seems analogous to mapping unfamiliar terrain. In this metaphor of a geography containing a number of cities, statistical difference tests allowed us to determine if the altitude of two cities differed, providing that we did not make this comparison between too many cities in one region. Our more sophisticated tools allowed us to describe a route from one city to another, but the more roads we described between two places, the less confidence we could have that we knew any reliable routes. It seems fundamentally flawed that the resolution of the primary quantitative analysis approaches widely understood in rehabilitation is so limited by the need to contain the risk of false positives. Measuring just one or two points on a landscape can give us almost no confidence that we truly understand the geography in which we are operating. In topographic mapping, the more data points you have, the better you can understand the terrain.

Quantitative modeling and analysis tools that bear more resemblance to topographic mapping would enable us to examine and learn from our data in great depth using techniques where closer investigation strengthens rather than weakens the interpretations we can draw from our data. There will no doubt be researchers who have been working in these areas for years, and it is time for their voices to be widely heard. Identification of approaches along these lines that already exist in the statistical literature and dissemination to make them accessible to rehabilitation researchers is a high priority.

Closing the feedback loop

Control of even highly overlearned behavior like speech relies fundamentally on continuous feedback loops. Disruption of such

feedback with interference (eg, presenting the average speaker with their own voice on a one fifth of a second delay) results in immediate and marked deterioration in speech performance.¹² It is reasonable to consider the importance of feedback loops in other, more complex behaviors—something that is, for instance, already having positive effect in mental health research.¹³ In rehabilitation practice, the primary goal is long-term adjustment of the person receiving rehabilitation—their level of community integration, meaningful activity to engage in, having close and warm personal relationships a positive adjustment to life. In most rehabilitation services, however, we only have limited feedback on individual outcomes at 6 months post-discharge and almost no feedback loop at all regarding longer-term outcomes for the people receiving our services.

Why do we have such limited long-term outcome data? As we increasingly deploy Rasch methods/item response theory methods, we are finally learning to properly measure the constructs we are interested in,^{14,15} but we still have an ongoing lack of certainty about how to properly define a good long-term outcome.¹⁶ Second, routinely collecting long-term outcome data has been logistically and financially impractical to undertake to date. While the definition of a good outcome may remain open for debate, the lack of data might, in the future, no longer be a constraint. As data can increasingly be automatically collected through ambient sensors¹¹ and (further in the future) sifted and refined with the assistance of artificial intelligence systems, longitudinal outcome data may become routinely available. This may not even be restricted to quantitative data. Longitudinal qualitative data^{17,18} are currently rare¹⁹ but can be illuminatory.

As the sheer breadth and depth of quantitative data increases by orders of magnitude in the future, and as researchers begin to have tools that enable them to collect genuinely qualitative data on scale, we may even find our methods start to meet in the middle. Well before that, it is worth considering that it may be time to leave such methodology turf wars behind. Increasing numbers of researchers are already reaping the benefits of mixed-methods research that combines quantitative and qualitative approaches.²⁰ We need the landscape paintings that qualitative research provides, which uniquely convey a rich depth of the rehabilitation experience not otherwise accessible to us. We also need the analysis equivalent of both photographic techniques and the precise accuracy of topographic maps: all three provide unique value through unique perspectives on the same rehabilitation landscape.

Computer-aided analysis

It is not unreasonable to expect that our analyses of rehabilitation data will be supported in the future by tools that take far more initiative. Instead of software packages that blindly run analyses we select on the tabulated data we provide, we will have tools that actively seek out clarification regarding the nature of our data, use that context to understand the data in ways we have not, and present to us views and analyses of our datasets that we have not anticipated. Sufficiently large datasets, collected automatically from clinical contexts, may be made possible by automated mechanisms for recording, coding, and filtering fine-grained data about rehabilitation processes. Tools that then connect this data to descriptions we provide of the constructs and real-world relationships would enable us to draw together a breadth of data that we might struggle to comprehend with our current approaches. It may be that apparently disparate parts of an interdisciplinary rehabilitation process interact in ways we have not understood. For example, we may discover that one therapeutic

Download English Version:

<https://daneshyari.com/en/article/6149542>

Download Persian Version:

<https://daneshyari.com/article/6149542>

[Daneshyari.com](https://daneshyari.com)