

Let me count that again

Roderick J McClure 

How much data are enough? How accurate do they have to be before they are useful? Do data have to be collected from us for them to be relevant to me?

Well, it depends, on a lot of things, but mostly on the questions we are hoping the data will help us address. And even when we have enough data, and they are sufficiently accurate, and they are relevant to us, they are still not much help unless they are the right data and we know how to use them.

Before unpacking these issues, let us first anchor our discussion in the roots of the discipline that underpinned the Journal's establishment: the epidemiological approach to injury prevention. Descriptive epidemiology is the science that grounds public health. As practitioners we cannot address what cannot 'see'. Descriptive epidemiology is the means by which we can elucidate the nature and extent of a problem and describe its distribution by time, person, place, severity, activity, location and mechanism. We can use descriptive data about the burden, opportunity and cost to prioritise our response. Data about availability and quality of preventive services and data about programme process, impact and outcome can help improve the performance of these preventive services. Data can be used to generate hypotheses about the cause, identify risk factors, quantify countermeasure efficacy and determine the effectiveness of programme implementation.¹

So how much data are enough? A more useful way of asking this question is 'How much data do I need before I can confidently act?' In 1964, Dr Terry, Surgeon General of the US Public Health Service, named cigarette smoking a cause of lung cancer and laryngeal cancer in men, a probable cause of lung cancer in women and the most important cause of chronic bronchitis.² How many years after that were vested interests still arguing the need for more data before those claims could be substantiated.³ It would be interesting

Correspondence to Dr Roderick J McClure, School of Rural Medicine, University of New England, Armidale, ME 04005, Australia; rmclure@une.edu.au

to examine in the same light many of the calls for more data on a range of contemporary issues.

How accurate do data need to be? This is similar to the 'how much?' question and has the same answer. Data need to be as accurate as they need to be to support confident action. If the work expended to achieve greater and greater data precision is not matched by equal precision of intervention, then the benefits of the increased accuracy may be lost. This is not to say actions should not be evidence based, but the importance of data limitations cannot be judged outside of the context within which data are being used.

The how much/how accurately questions frequently arise together in conversations about the use of routine surveillance data to address post hoc research questions. Routine surveillance data and research data that have been prospectively collected to address a properly formulated research question are as different as drinking tea and chopping wood. To complain that one is not the other makes little sense. It also makes little sense to criticise generic surveillance data because they are not as specific for a given purpose as a purpose-built injury-specific database because to redress that problem would create more databases than society could afford to maintain.

Do we need local data for local implementation? Yes, but generally not because causal relationships differ between locations, but because local data create a local legitimacy for the sometimes inconvenient interventions.

We need enough data, they need to be accurate and they need to be local, but above all, they need to be useful. There can be data, data everywhere yet still not sufficient information to make a difference. Prevalence, incidence rates and risk ratio statistics lack the vitality of the real-time analytics that drive business in large companies. Static models of historical correlations lack the decision support capability of dynamic models. Rarely do datasets contain the social-level variables that are critical determinants of effective

implementation. As researchers and practitioners, it is our responsibility to improve data quantity and quality, but we should also focus on collecting what we need and only what we need to achieve our preventive goals.

I suggest you turn to the education section to this issue and see what some of our most experienced colleagues have to say on the subject,⁴⁻⁷ and then try and answer this question. If you were given a one million dollar grant to address your injury issue of concern, would you spend it on collecting data or on implementing a prevention programme?

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; internally peer reviewed.

© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.



To cite McClure RJ. *Inj Prev* 2020;**26**:301.

Inj Prev 2020;**26**:301.
doi:10.1136/injuryprev-2020-043906

ORCID iD

Roderick J McClure <http://orcid.org/0000-0002-9067-8282>

REFERENCES

- 1 McClure RJ, Mack K. Injury surveillance as a distributed system of systems. *Inj Prev* 2016;**22**:i1–2.
- 2 US Department of Health, Education, and Welfare. *Smoking and health: report of the Advisory Committee to the surgeon General of the public health service*. pHS publication No. 1103. Washington: U.S. Department of Health, Education, and Welfare, Public Health Service, Center for Disease Control, 1964.
- 3 Milberger S, Davis RM, Douglas CE, et al. Tobacco manufacturers' defence against plaintiffs' claims of cancer causation: throwing mud at the wall and hoping some of it will stick. *Tob Control* 2006;**15**:iv17–26.
- 4 Holder Y. Data capture for injury prevention. *Inj Prev* 2020;**26**:391–4.
- 5 Lyons RA. How much does quality matter: the value of data. *Inj Prev* 2019;**26**:397–9.
- 6 Hemenway D. Let's make it a priority to improve injury data. *Inj Prev* 2020;**26**:395–6.
- 7 Vallmuur K. Artificial intelligence or manufactured stupidity? The need for injury informaticians in the big data era. *Inj Prev* 2019;**26**:400–1.