Issues with suicide databases in forensic research

Roger W. Byard¹

Accepted: 21 February 2017 / Published online: 9 March 2017
© Springer Science+Business Media New York 2017

Things are seldom what they seem
William Gilbert 1836-1911

Although it is well recognized in computing circles that poor quality input results in poor quality output, this is sometimes not appreciated in areas of research that rely on computerized databases for providing information on trends in injury and disease, and in informing educational and prevention campaigns. A case in point is the paper by Austin et al. in this issue of the journal. The group analyzed local suicide data from South Australia and compared it to data from the same population held in national databases, to see if there were similarities or differences [1].

The study clearly demonstrated that there were significant differences in rates of suicide depending on whether local data were evaluated, or the figures were taken from either of the two national databases (The National Coronial Information System – NCIS, or the Australian Bureau of Statistics – ABS). Specifically, the suicide rate in South Australia was listed as 13.3 per 100,000 based on evaluation of local data, and 12.4 and 12.3 on ABS and NCIS databases. The biggest discrepancy occurred with drug overdose suicides, with only 67.8% recorded on the NCIS database [1]. This creates difficulties if trends are to be monitored and decisions made based on data that may not accurately reflect what is going on in the communities served.

Issues with the reliability of data on suicide are not new, with some studies demonstrating a fourfold difference between the official rates and those that were calculated based on reclassifications [2]. It is suggested that problems occur when there are changes in rules for classifying deaths or for collecting mortality data, and when alterations occur in diagnostic methods and in medical terminology [2]. Although a high percentage of cases of suicide were confirmed in a Nordic study, 9% of the Norwegian cases of accidental and natural deaths were changed to suicide after reclassification, as were 21% of cases initially considered to be “undetermined” in the Swedish cohort [3]. Other problems in general data base entry occur when there is misinterpretation of primary information or simply when mistakes are made with the initial entry [4].

A classic example of how alterations in medical terminology can influence official mortality data occurred in the 1960’s when the rate of sudden infant death syndrome increased dramatically in a number of countries due solely to a diagnostic shift by pathologists who had moved away from assigning the causes of these deaths to respiratory infections. The possibility of diagnostic transfer should always be considered when mortality rates appear to be changing [5].

An Australian study comparing the data from the Queensland Suicide Register with the Australian Bureau of Statistics showed statistically significant underestimation of suicide numbers in that state by the Bureau in 24 of 28 pair-wise comparisons. A recount of national suicides in Australia recorded by the Bureau in 2004 showed an underestimation of 16% [6]. It has been suggested that the Spanish Statistical Office (Instituto Nacional de Estadística, INE) underestimated suicides in that country by 443.86 cases per year between 2006 and 2010 [7].

It is of course not only suicide data that present challenges. Examination of the results derived from different databases of

✉ Roger W. Byard
roger.byard@sa.gov.au

¹ School of Medicine, The University of Adelaide, Frome Road, Level 3 Medical School North Building, Adelaide, SA 5005, Australia

patients suffering from acute renal failure in the United States has shown variations in incidence from 0.9 to 20% and in mortality from 25 to 80%. The possibility of bias introduced by using different definitions of acute renal failure at different times was suggested as one explanation for these discrepancies [8]. Another American study published in 2007 revealed “major, and progressively increasing, discrepancies between two U.S. federal databases” that tabulated critical care unit and hospital usage and Medicare costs. The differences were thought to be due to differences in the types of codes that had been included [9].

Although it has been proposed that official rates of suicide can still be useful in comparing trends and features among cultural and social groups because the sources of error are random [10], other researchers have found that many of the errors in clinical databases were in fact non-random [4]. It was pointed out that the errors occurred in “special and cognitive clusters” that could “potentially affect the interpretation of the study results” [4]. For example, idiosyncratic coding by one person entering data could result in a major skew in numbers of cases in specific categories.

What does this all mean for forensic research? While national databases can certainly provide useful information, it obviously must be handled circumspectly with a realization that capture of data may not be complete (i.e. neither national nor for individual states), and that total category numbers may not accurately reflect true community data. It is imperative, therefore, that researchers clearly state the limitations of their data in any subsequent publications. Perhaps more credence should be placed on smaller case series and local studies, where all of the cases have been reviewed and if necessary reclassified by researchers expert in the field [11, 12], and not by administrative assistants who may have no particular expertise with, or interest in, the data that they are handling. In this way significant local trends that may otherwise be obscured by national figures may be far more easily identified, with the confidence that as much accurate information as

possible has been gathered about the cases: i.e. to paraphrase a popular saying: high quality in-high quality out (HIHO).

References

1. Austin A, van den Heuvel C, Byard RW. Differences in local and national database recordings of deaths from suicide. *Forensic Sci Med Pathol*. doi:10.1007/s12024-017-9853-x.s.
2. Tøllefsen IM, Thiblin I, Helweg-Larsen K, Hem E, Kastrup M, Nyberg U, et al. Accidents and undetermined deaths: re-evaluation of nationwide samples from the Scandinavian countries. *BMC Pub Health*. 2016;16:449.
3. Tøllefsen IM, Helweg-Larsen K, Thiblin I, Hem E, Kastrup M, Nyberg U, et al. Are suicide deaths under-reported? Nationwide re-evaluations of 1800 deaths in Scandinavia. *BMJ Open*. 2015;5:e009120.
4. Goldberg SI, Niemierko A, Turchin A. Analysis of data errors in clinical research databases. *AMIA 2008 Symposium Proceedings*: 242–6.
5. Mitchell E, Krous HF, Donald T, Byard RW. Changing trends in the diagnosis of sudden infant death. *Am J Forensic Med Pathol*. 2000;21:311–4.
6. Williams RF, Doessel DP, Svetcic J, De Leo D. Accuracy of official suicide mortality data in Queensland. *Aust N Z J Psychiatry*. 2010;44:815–22.
7. Giner L, Guija JA. Number of suicides in Spain: differences between data from the Spanish statistical office and the Institutes of Legal Medicine. *Rev Psiquiatr Salud Ment (Barc)*. 2014;7:139–46.
8. Lameire N, Van Biesen W, Vanholder R. The rise of prevalence and the fall of mortality of patients with acute renal failure: what the analysis of two databases does and does not tell us. *J Am Soc Nephrol*. 2006;17:923–5.
9. Halpern NA, Pastores SM, Thaler HT, Greenstein RJ. Critical care medicine use and cost among Medicare beneficiaries 1995–2000: major discrepancies between two United States federal Medicare databases. *Crit Care Med*. 2007;35:692–9.
10. Sainsbury P. Validity and reliability of trends in suicide statistics. <http://europepmc.org/abstract/med/6678086>. Accessed 18 Feb 2017.
11. Austin A, van den Heuvel C, Byard RW. Causes of community suicides among indigenous south Australians. *J Forensic Legal Med*. 2011;18:299–301.
12. Austin A, Byard RW. Prison suicides in South Australia 1996–2010. *J Forensic Sci*. 2014;59:1260–2.