

EDITORIAL



Is research from databases reliable? Not sure

Meri R. J. Varkila^{1,2*} and Olaf L. Cremer²

© 2018 Springer-Verlag GmbH Germany, part of Springer Nature

Introduction

As health services become increasingly digitalized, the volume of clinical data being collected has reached immense proportions globally. However, much of this information has been primarily documented for purposes other than research, and is accessible only through secondary sources, such as administrative, quality assurance, or national audit databases. Although such data sources frequently provide scientists with access to information about many thousands of subjects and thus may offer unprecedented opportunities for research, they can present significant challenges when it comes to validity, granularity, and reproducibility of data. Thus, what may first seem like a treasure trove may prove to come at significant costs.

Randomization is not the only road to causal inference

Although randomized controlled trials (RCT) are still considered superior in the general hierarchy of study designs, randomized interventions may sometimes be simply unfeasible. Furthermore, RCTs frequently have yielded inconclusive or conflicting results when performed in ICU settings [1]. Most likely this is related to the large heterogeneity and complexity of disease processes that are intrinsic to critically ill patients [2]. Observational studies exploiting existing data may offer a worthy alternative in this respect.

Comparative evaluations of observational studies and RCTs assessing the same interventions have reported inconsistent results, with some meta-analyses yielding only small, yet others very substantial differences in effect estimates [3, 4]. However, the fact that the smallest discrepancies occurred in observational studies that were estimated to have lower risk of bias indicates that such deviations can be prevented (at least in part) by using adequate methodology. Thus, the value of well-designed observational studies employing correct analytical techniques and high-quality data should not be underestimated. This holds particularly true for research questions evaluating existing medical practices that are (almost) universally accepted, for uncovering adverse effects of treatment, and in situations of rare disease. Database studies may generate new hypotheses, and help justify interventional trials regarding issues where no formal assessment yet exists. The observational study by Connors et al. [5], questioning the use of pulmonary artery catheterization in the management of critically ill patients, is a classic example of a study that revealed unexpected results using existing data, and consequently led to critical reappraisal of a common practice (even though suggested harm was not confirmed by experimental studies) [6]. Observational studies using administrative data can also offer insights in situations where an intervention simply cannot be randomly allocated in any practical way, e.g., in studying the (potentially harmful) effects of premature discharge at night [7].

Quality before quantity

Despite the potential utility of observational research, confounding by indication and information bias pose well-known challenges in observational studies in general, and in administrative database research in particular. As decision-making in the ICU is frequently based on imprecise (or even subjective) clinical

*Correspondence: m.r.j.varkila@umcutrecht.nl

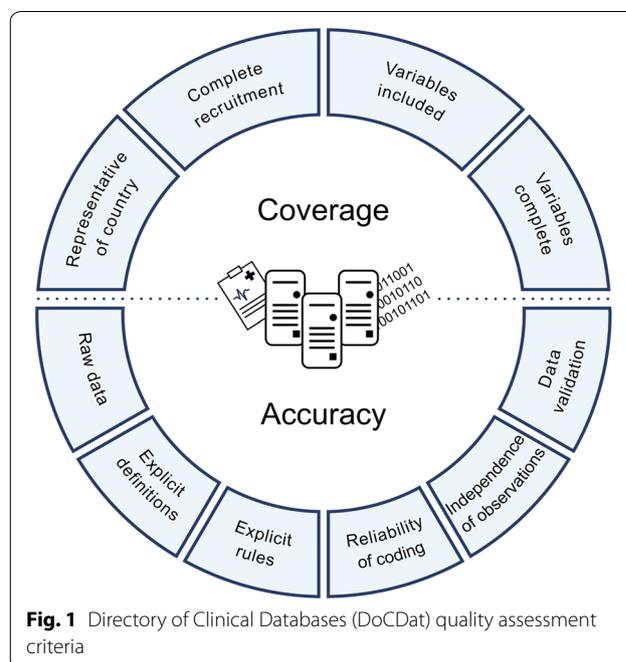
¹ Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands
Full author information is available at the end of the article

For contrasting viewpoints, please go to doi: <https://doi.org/10.1007/s00134-018-5073-4> and doi: <https://doi.org/10.1007/s00134-018-5436-x>.

criteria—which are typically not well captured by the data—factors related to the exposure and/or outcome may remain unobserved. Furthermore, the information contained in large administrative databases often lacks the temporal resolution necessary to adequately model time-dependent exposures, adjust for immortal time biases, and account for competing events. Databases may also contain (non-random) missing data. Together, this renders database research susceptible to spurious findings. In this respect it is important to note that, as sample size grows, so does a study's capacity to attain a statistically significant result, regardless of whether the observed associations are true or not [8]. Robust statistical methodologies and a good deal of common sense are therefore required during analysis.

Another potential pitfall when using administrative databases for research relates to (hidden) incentives for information capture, as secular trends in coding criteria and reimbursement policies for hospital claims can have a significant impact on data registration. This is nicely illustrated by a study analyzing hospital claims data over a 7-year period [9]. Following revised *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9) coding instructions and a new reimbursement scheme by the Centers for Medicare and Medicaid Services (CMS) in 2003, US hospitalization rates for patients with a principal diagnosis of pneumonia seemingly decreased, whereas sepsis diagnoses nearly tripled. A similar effect was seen after CMS reimbursement for sepsis care became coupled to severity of illness in 2007 [10]. However, when claims-based methods were compared to more precise criteria, true sepsis incidence actually appeared to remain stable [11].

Even when data capture is relatively complete and consistent over time, not all databases are created equal. The quality and suitability of any data source therefore need to be appraised considering the context of the specific research question at hand. To aid investigators in this process, the Directory of Clinical Databases (DoCDat) Development Group has developed a framework for assessing database quality across two methodological dimensions—coverage and accuracy (Fig. 1) [12]. Coverage refers to (1) the extent to which all eligible study subjects have been included, as well as (2) the representativeness, completeness, and scope of collected data. Accuracy is determined by (1) the form and extent in which (quantitative) data is collected, coded, and subsequently validated, (2) the independence of observations in relation to study outcomes, and (3) the use of explicit variable definitions and rules for data collection. In addition to informing rational selection of potential databases for research, a systematic appraisal of data sources using



standardized methodology can help identify and address points of qualitative concern and improve study validity.

Future perspectives and final thoughts

While explicit reporting and even prior publication of study protocols and statistical analysis plans have become standard practice when performing clinical trials, (administrative) database research has generally remained exempt from such requirements. However, even though observational studies still remain susceptible to *p*-hacking and fishing expeditions, standard operating procedures, improved reporting of data collection methods, and the use of more precise variable definitions could help remedy some of the issues related to administrative database research [13].

In addition, defining exposures and outcomes based on clinical data contained in electronic health records has shown higher sensitivity and better consistency than is possible on the basis of ICD-9 and similar diagnostic or procedural codes contained in administrative databases [11, 14]. Thus, the quality of database research could possibly be improved by the introduction of (semi)automated data surveillance algorithms utilizing raw clinical information for robust capture and parameterization of study variables. In the ICU, where bedside observations and health records have already become largely digitized, smart data mining seems an increasingly feasible endeavor.

As data capture becomes more advanced, so will study designs and methods for data analysis. An interesting future prospect that is both linked to the domains of machine-learning and Bayesian statistics is the fusion of

an interventional and observational study design into a single *randomized, embedded, multifactorial, adaptive platform (REMAP) trial*. By implementing response-adaptive randomization based on data that are routinely collected in the electronic health record, REMAP trials can identify best-performing treatment regimens while providing continuous quality improvement for patients [15].

In conclusion, large administrative databases offer excellent opportunities for observational research. Besides their indisputable value for diagnostic, prognostic, and descriptive research, administrative databases containing large patient samples and data collected over prolonged periods of time can also be used to investigate specific etiological hypotheses. However, when working with administrative databases, transparent reporting, rigorous assessment of data quality, thoughtful statistical analysis, and critical questioning of *why* and *how* data were collected all remain essential.

Author details

¹ Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands. ² Department of Intensive Care Medicine, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands.

Compliance with ethical standards

Conflicts of interest

The authors report no conflicts of interest.

Received: 27 November 2018 Accepted: 7 December 2018

Published online: 14 December 2018

References

- Ospina-Tascón GA, Büchele GL, Vincent JL (2008) Multicenter, randomized, controlled trials evaluating mortality in intensive care: doomed to fail? *Crit Care Med* 36(4):1311–1322
- de Grooth HJ, Postema J, Loer SA, Parienti JJ, Oudemans-van Straaten HM, Girbes AR (2018) Unexplained mortality differences between septic shock trials: a systematic analysis of population characteristics and control-group mortality rates. *Intensive Care Med* 44(3):311–322
- Anglemyer A, Horvath HT, Bero L (2014) Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev* 4:MR000034. <https://doi.org/10.1002/14651858.MR000034.pub2>
- Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP (2016) Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. *BMJ* 352:i493. <https://doi.org/10.1136/bmj.i493>
- Connors AF Jr, Speroff T, Dawson NV, Thomas C, Harrell FE Jr, Wagner D, Desbiens N, Goldman L, Wu AW, Califf RM, Fulkerson WJ Jr, Vidaillet H, Broste S, Bellamy P, Lynn J, Knaus WA (1996) The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. *JAMA* 276(11):889–897
- Rajaram SS, Desai NK, Kalra A, Gajera M, Cavanaugh SK, Brampton W, Young D, Harvey S, Rowan K (2013) Pulmonary artery catheters for adult patients in intensive care. *Cochrane Database Syst Rev* 2:CD003408. <https://doi.org/10.1002/14651858.CD003408.pub3>
- Vollam S, Dutton S, Lamb S, Petrinic T, Young JD, Watkinson P (2018) Out-of-hours discharge from intensive care, in-hospital mortality and intensive care readmission rates: a systematic review and meta-analysis. *Intensive Care Med* 44(7):1115–1129. <https://doi.org/10.1007/s00134-018-5245-2>
- van Walraven C, Austin P (2012) Administrative database research has unique characteristics that can risk biased results. *J Clin Epidemiol* 65(2):126–131
- Lindenauer PK, Lagu T, Shieh M-S, Pekow PS, Rothberg MB (2012) Association of diagnostic coding with trends in hospitalizations and mortality of patients with pneumonia, 2003–2009. *JAMA* 307(13):1405–1413
- Gohil SK, Cao C, Phelan M, Tjoa T, Rhee C, Platt R, Huang SS (2016) Impact of policies on the rise in sepsis incidence, 2000–2010. *Clin Infect Dis* 62(6):695–703
- Rhee C, Dantes R, Epstein L, Murphy DJ, Seymour CW, Iwashyna TJ, Kadri SS, Angus DC, Danner RL, Fiore AE, Jernigan JA, Martin GS, Septimus E, Warren DK, Karcz A, Chan C, Menchaca JT, Wang R, Gruber S, Klompas M, CDC Prevention Epicenter Program (2017) Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009–2014. *JAMA* 318(13):1241–1249
- Black N, Payne M (2003) Directory of clinical databases: improving and promoting their use. *BMJ Qual Saf* 12:348–352
- Kaufmann T, Granholm A, Keus F, Myburgh J, Perner A, Möller MH, van der Horst ICC (2018) Foresight over hindsight: mandatory publication of clinical research protocols prior to conduct. *Acta Anaesthesiol Scand* 35:85. <https://doi.org/10.1111/aas.13267>
- Rhee C, Jentzsch MS, Kadri SS, Seymour CW, Angus DC, Murphy DJ, Martin GS, Dantes RB, Epstein L, Fiore AE, Jernigan JA, Danner RL, Warren DK, Septimus EJ, Hickok J, Poland RE, Jin R, Fram D, Schaaf R, Wang R, Klompas M, Centers for Disease Control and Prevention (CDC) Prevention Epicenters Program (2018) Variation in identifying sepsis and organ dysfunction using administrative versus electronic clinical data and impact on hospital outcome comparisons. *Crit Care Med* 45:85. <https://doi.org/10.1097/ccm.0000000000003554>
- Angus DC (2015) Fusing randomized trials with big data: the key to self-learning health care systems? *JAMA* 314(8):767–768