ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 699–701

Journal of Clinical Epidemiology

EDITORIAL

Has the time arrived for clinical epidemiologists to routinely use 'routinely collected data'?

The comparative effectiveness movement has rightly concluded that experimental designs can only address one segment of the most important questions and that we just have to find robust ways to use routinely collected data. In this issue, Benchimol et al. outline in a commentary the importance of developing guidelines for the Reporting of studies Conducted using Routinely collected Data (RECORD) statement. This includes guidelines for use of health administrative data, data warehouses of electronic medical records, primary care medical record data, and disease registries. The RECORD guidelines will be developed in close conjunction with members of the STROBE group to ensure consistent methods and to make this a useful addition to the original STROBE statement. This process will involve general consultation with stakeholders, a formal Delphi process, and the eventual production of useful and widely accepted reporting guidelines. Details for joining the process are included in the article should *JCE* readers be interested in joining the process as a stakeholder.

In 1997 Bain et al. published an article entitled ''Routinely collected data in national and regional databases-an underused resource'' [1]. They argued that routinely collected data may be used in a number of different ways beyond their traditional uses for population health assessments and health service planning; the strengths of the best routinely collected databases are their population coverage, large number of subjects, long duration, and their low cost. They have been used by a minority of clinical epidemiologists because of concerns about quality and the absence of clinical data on aspects such a clinical severity. Impressive advances on both these fronts have been made. The time has come for this to be reassessed, and indeed now there are ongoing constructive discussions between clinical epidemiologists, pharmacoepidemiologists, health services, public health, and health systems researchers within the Cochrane Collaboration. Fortunately, there have been major advances in establishing high-quality representative administrative databases that can be linked with clinical registries and other databases to evaluate the impact of important aspects of clinical care, public health, and health systems. One such is the Institute of Clinical Evaluation Studies (ICES) in Ontario, which is celebrating its 20th anniversary this year (http://www.ices.on.ca/file/TWENTY_ICES_20th_Anniversary_Magazine.pdf). This institute has administrative data on all the interactions within the health system of all 13 million individuals in Ontario; this is supplemented by registries such as the Canadian Stroke Registry, Implantable Cardiac Device Registry, and the Ontario Biologics Registry. Notable examples relevant to clinical epidemiology and evidence-based health care impacts have been the effects of publication of introducing wait times algorithms for cardiac surgery patient safety such as the higher mortality seen in hospitals; atlases on chronic disease showing the local prevalence rates that are needed for pretest probabilities and for applying GRADE (http://www.gradeworkinggroup.org/) to locally appropriate recommendations; atlases of variations in rates of events and practices to provide peer pressure to improve practices (e.g., appendectomy rates, caesarian rates, c.difficile hospital infections, cardiac events in hospitals; dangers of cell phones while driving; diabetes rates in vulnerable groups and their access to care and nutritious food; the finding that spironolactone was associated with toxic levels of potassium; and atypical bone fractures with bisphosphonates).

In a related vein, in his provocatively entitled Commentary ''Quality-of-life data should be regarded as a vital sign,' Feeny argues that the health care system systematically collects data on inputs, such as risk factors, processes of care, and the amount and type of services provided, but the health system only sporadically collects data on the health status and health-related quality of life (HRQL) of those served. He argues that health-related quality-of-life measures should be an integral component of chronic care management and be routinely used to monitor the performance of the health care system. Patients and clinicians will need to be trained to exploit this powerful source of information, and administrators will need to support those activities.

Quality of life. Tan et al. demonstrated measurement equivalence of the English and Chinese versions of the Short Form 36 version 2 (SF-36v2) and Short Form 6D (SF-6D). Thus, because the two languages measure the same construct, the results can be pooled together and conclusions can be made to both the language-speaking group as a whole.

Instrument overload is always a concern, but should we accept that a short form of even a widely used complete original questionnaire inherits the same psychometric

0895-4356/$ - see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jclinepi.2013.04.004

properties and is equally valid and responsive? It may well be that the improved feasibility may come at the cost of worse validity (e.g., content) and responsiveness. Goetz et al. found major worrying psychometric gaps in some widely used questionnaires when they reviewed current practice and update guidelines for the methodology of shortening composite measurement scales. The authors propose some criteria for item reduction based on a rigorous methodology is necessary if the short-form instrument aims to maintain the validity and other measurement properties of the parent instrument, which in turn supports application in research and clinical practice. Consensus and a dissemination plan are needed if this parlous state is to change.

The next installment of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) series is published this month with 2 papers by Andrews et al. on classifying the direction and strength of recommendations plus a report on the importance of training to minimize interrater reliability. The first article posits that the strength of a recommendation, separated into strong and weak, is defined as the extent to which one can be confident that the desirable effects of an intervention outweigh its undesirable effects. The second article provides guidance on how to move from the Grade Quality of Evidence to the Grade Recommendations. This latter is to make explicit the basis of the Recommendation that is essentially the balance of desirable and undesirable outcomes of interest among alternative management strategies; this depends on four domains, namely estimates of effect for desirable and undesirable outcomes of interest, confidence in the estimates of effect, estimates of values and preferences, and resource use. This article will be of notable interest to those readers who have been awaiting more detailed guidance on the values and preferences of patients and individuals affected by or paying for the care of the condition of interest. We agree that panels should strive to provide recommendations when asked; tough decisions should not by avoided by taking the often all-to-easy way of declining to make recommendations and asking for more research and support, which is the case made in these articles. In the third article, by Mustafa et al., the inter-rater reliability of assessing the quality of evidence using the GRADE approach was evaluated. Their findings suggest that individuals trained using the GRADE approach improve reliability in comparison to intuitive judgments about the quality of evidence and that 2 individual raters can reliably assess the quality of evidence using the GRADE system.

Statistics. When the number of events is small during Cox regression analysis because the events are rare or because it is a cohort study with long latency events, it is unclear what alternative analytical strategies can be used and when this type of alternative approach is needed. A study by Lin et al. explores such alternative analytical strategies. They found that when the primary predictor is categorical and when the events per variable is less than 6, the Firth method can be used to obtain more accurate parameter estimates. Appropriate choices of priors using Bayesian analysis should also be able to increase the accuracy of parameter estimates. The drawback of this approach, however, is that it requires more expertise. In another article, Workman and colleagues provide a sample size formula for cluster randomized stepped wedge designs. The stepped wedge design is increasingly being used in cluster randomized trials. However, there is little information available about the design and analysis strategies for these kinds of trials. Sample sizes are smaller than with other designs, but a simple sample size formula has been lacking. The formula they present in this paper corrects for clustering as well as for the design.

Clinician compliance is as important as patient adherence in achieving improved health outcomes. Little is known about the relationship between clinical experience and clinician compliance with the study protocol in randomized clinical trials. Walter et al., in the context of a recent randomized trial of surgical techniques for tibia fracture fixation, raise the hypothesis that varying amounts of clinical experience has been previously underemphasised as an important cause of noncompliance by clinicians with the randomly assigned treatment that can be highly detrimental to a study's power may. They found that, in randomized clinical trials, noncompliance by clinicians with the randomly assigned treatment was detrimental to the power of the study. They conclude that further research is needed in this area to identify, quantify, and understand the factors associated with noncompliance, including clinical experience.

Quality of life. Should we systematically throw out widely used existing scales for measuring Patient Reported Outcomes (PROs) if they fail the requirements of the approval agencies, such as the FDA guidance documents [2]? Alliet et al. sought to determine the content, structural, and construct validity of the Dutch version of the Neck Disability Index (NDI). They found that in addition to a suboptimal translation, there was a lack of consensus on the construct the NDI intends to measure. The authors advocate the development of a new disease-specific instrument, starting from a clear definition of the construct to be measured and using more advanced psychometric techniques. In the same spirit of patient-centeredness, Dannell and colleagues sought to examine the difference between children and their parents in reporting symptoms and treatment of allergic diseases within a longitudinal birth cohort. Although reports of allergic symptoms and treatment by 12-year-old children and their parents were in moderate to good agreement, children reported more symptoms than their parents. The authors conclude that symptoms of allergic disease should be reported by children themselves, from the age of 11 years, whereas questions of prescribed pharmacologic treatment could be answered either by the children or their parents.

Scoring systems for quality-of-life instruments need to trade off their discriminative ability with simplicity.

A paper by Forero et al. proposes a multidimensional item response theory (MIRT) scoring system for the Short Form 12 (SF-12). The MIRT models showed a clear construct structure for the physical and mental component score dimensions. Results support the use of SF-12 MIRT-based scores as a valid and reliable option to assess health status, but the score still needs computation rather simple addition.

Finally, the third paper in our series on effective writing and publishing of scientific papers provides some tips for structuring and writing a concise introduction. Cals and Kotz introduce the concept of the 'funnel,' starting with the broad context and ending with a well-focused article. Please read the article for more details.

Peter Tugwell
J. André Knottnerus
Leanne Idzerda
Editors
E-mail address: lidzerda@uottawa.ca (L. Idzerda)

References

[1] Bain MR, Chalmers JW, Brewster DH. Routinely collected data in national and regional databases- an under-used resource. J Public Health Med 1997;19:413−8.

[2] US Food and Drug Administration. Patient-reported outcome measures: use in medical product development to support labeling claims. Rockville, MD: Department of Health and Human Services, US Food and Drug Administration, Center for Drug Evaluation and Research; 2009.