Journal of Clinical Epidemiology 82 (2017) 137−148

Journal of Clinical Epidemiology

Four hundred or more participants needed for stable contingency table estimates of clinical prediction rule performance

Peter Kent[a,b,*], Eleanor Boyle[b,c], Jennifer L. Keating[d], Hanne B. Albert[e], Jan Hartvigsen[b,f]

[a]*School of Physiotherapy and Exercise Science, Curtin University, Kent Street, Bently, Perth, Western Australia 6102, Australia*

[b]*Clinical Biomechanics Research Unit, Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Campusvej 55, Odense M 5230, Denmark*

[c]*Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, 155 College St, Toronto, Ontario, ON M5T 3M7, Canada*

[d]*Department of Physiotherapy, Faculty of Medicine Nursing and Health Sciences, Monash University, McMahons Road Frankston, Melbourne, VIC 3199, Australia*

[e]*The Modic Clinic, Jernbanegade 4, Odense 5000, Denmark*

[f]*Nordic Institute of Chiropractic and Clinical Biomechanics, Campusvej 55, University of Southern Denmark, Odense M 5230, Denmark*

Accepted 11 October 2016; Published online 12 November 2016

Abstract

Objectives: To quantify variability in the results of statistical analyses based on contingency tables and discuss the implications for the choice of sample size for studies that derive clinical prediction rules.

Study Design and Setting: An analysis of three pre-existing sets of large cohort data (n = 4,062−8,674) was performed. In each data set, repeated random sampling of various sample sizes, from n = 100 up to n = 2,000, was performed 100 times at each sample size and the variability in estimates of sensitivity, specificity, positive and negative likelihood ratios, posttest probabilities, odds ratios, and risk/prevalence ratios for each sample size was calculated.

Results: There were very wide, and statistically significant, differences in estimates derived from contingency tables from the same data set when calculated in sample sizes below 400 people, and typically, this variability stabilized in samples of 400−600 people. Although estimates of prevalence also varied significantly in samples below 600 people, that relationship only explains a small component of the variability in these statistical parameters.

Conclusion: To reduce sample-specific variability, contingency tables should consist of 400 participants or more when used to derive clinical prediction rules or test their performance.

Keywords: Clinical prediction rule; Sample size; Reproducibility of results; Epidemiologic research design; Predictive value of tests; Decision support techniques

1. Introduction

Clinical prediction rules are tools that define the relationship between multiple predictors (e.g., from an individual patient's history, physical examination, and/or test results) and likely diagnosis, prognosis, or treatment response [1,2]. They can be used to identify clinically relevant subgroups of patients. There is growing interest in clinical prediction rules, as seen in a recent study that identified more than 400 unique prediction rules across a range of health conditions that had been derived and published between 1965 and 2009, with the 80% of them published since the year 2000 [1].

Clinical prediction rules are derived from multivariable prediction models. The typical sequence is that candidate predictor variables are formed into prediction models using a variety of statistical methods, a final model is chosen based on its performance measures, and then that prediction model is transformed into a prediction rule [3]. Although the derivation of the rule from the model can also occur using a variety of statistical approaches, they often involve the

Funding: P.K. was partially funded by the Danish Fund for Chiropractic Research and Postgraduate Education. No funding source played any role in the scientific conduct of the study.

Conflict of interest: The authors declare there to be no conflicts of interest.

* Corresponding author. Tel.: (+61)-8-9266-3629; fax: (+61)-8-9266-3131.

E-mail address: peter.kent@curtin.edu.au (P. Kent).

http://dx.doi.org/10.1016/j.jclinepi.2016.10.004
0895-4356/

What is new?

Key findings

- There is a lack of information about appropriate sample sizes for studies that derive or test clinical prediction rules using contingency tables.
- We found very wide and statistically significant variability in estimates derived from contingency tables (sensitivity, specificity, positive and negative likelihood ratios, posttest probabilities, odds ratios, and risk/prevalence ratios) when calculated in sample sizes of 100 or 200 people, which typically stabilized in samples of 400−600 or more people.
- Although estimates of prevalence also varied significantly in samples below 600 people, in less than 15% of occasions was there less variability in samples extracted with a fixed prevalence than in samples with a varying prevalence.

What this adds to what was known?

- We are not aware of other studies that have investigated sample sizes requirements for studies that derive clinical prediction rules or measure prediction rule performance.

What is the implication and what should change now?

- Sample sizes in studies that derive prediction rules or measure prediction rule performance using contingency tables should consist of 400 participants or more.

use of statistics based on dichotomization of data into 2 × 2 contingency tables.

The 2 × 2 contingency table represents a dichotomized predictor variable and dichotomized outcome variable (the numbers of people who have/do not have a clinical characteristic present who also have/do not have a particular outcome). Dichotomized predictor and outcome variables in a contingency table enable the estimation of sensitivity, specificity, likelihood ratios, odds ratios, risk or prevalence ratios, and pretest and posttest probabilities. The clinical use of posttest probabilities is considered to be a high-level application of evidence-based care for the diagnosis of, and treatment selection for, individual patients [3].

Contingency tables have been used at various stages in the derivation of prediction rules. For example, in the case of the Flynn prediction for spinal manipulation in people with low back pain [4], univariate screening was initially used as a selection process to reduce the number of candidate variables, then continuous scale variables were dichotomized using the results of ROC analysis, and their sensitivity, specificity, and positive likelihood ratios were calculated from contingency tables for descriptive purposes, before the remaining candidate variables being entered into a logistic regression model. In other examples, contingency tables are used when identifying the number of items that need to be positive before a person is classified as ''rule positive'' or in measuring prediction rule performance [3]. Even when a prediction rule is created using some form of sum score from a multivariable model such as linear regression, simple dichotomization of ''over or under'' a threshold indicator and ''with or without'' the outcome of interest is often used in the process of rule calibration or for describing model performance. Similarly, recursive partitioning approaches to studying diagnostic pathways, such as Classification and Regression Trees, are based on contingency tables and provide predicted probabilities of a diagnosis [5]. So, the use of statistical estimates based on contingency tables commonly occurs at some stage in the creation of prediction rules, regardless of the overall method pathway used.

However, there is evidence that estimates based on contingency table statistics are highly variable across samples, due to variations in prevalence (selection bias) and disease severity (spectrum bias) [6−8]. These estimates can also be highly variable within samples, due to the presence of other clinical characteristics that may reflect the existence of subgroups in the sample [9,10]. Although the influence of these attributes (selection bias, spectrum bias, and the presence of clinical subgroups) on the variability in estimates based on contingency table statistics has been investigated [6−10], variability in estimates due to sample size has not been adequately researched.

Currently, the a priori estimation of adequate sample size is difficult in studies designed to derive clinical prediction rules, as (1) the performance characteristics of the rule cannot be known a priori and (2) the prevalence and severity of a particular health condition in a particular clinical setting may not be known. Sample sizes for studies that have derived musculoskeletal prediction rules have varied greatly, from 54 [11] to 8,924 [12], and are often less than 100 [4,11,13,14].

Therefore, the aims of this study were to (1) quantify variability in the estimates of clinical prediction rule performance (sensitivity, specificity, positive and negative likelihood ratios, posttest probabilities, odds ratios, and risk/prevalence ratios) that typically result from contingency tables of dichotomized predictors and outcomes and (2) discuss the implications of the results for sample sizes decisions in future studies.

2. Methods

2.1. Method summary

Three pre-existing sets of Danish cohort data were analyzed. The first data set was of 4,062 patients with spine