Children and Youth Services Review 56 (2015) 76–85

Contents lists available at ScienceDirect

Children and Youth Services Review

journal homepage: www.elsevier.com/locate/childyouth

Child abuse/neglect risk assessment under field practice conditions: Tests of external and temporal validity and comparison with heart disease prediction

Will Johnson [a,*], Thomas Clancy [a], Pascal Bastian [b]

[a] California State University, East Bay, Hayward, CA, USA
[b] Johannes Gutenberg University of Mainz, Mainz, Germany

ARTICLE INFO

Article history:
Received 2 April 2013
Received in revised form 23 June 2015
Accepted 23 June 2015
Available online 28 June 2015

Keywords:
Child abuse
Child neglect
Child maltreatment
Coronary heart disease
Risk assessment
Prognostic modeling
Prediction

ABSTRACT

Objectives: (1) Identify validation design and accuracy assessment standards for medical prognostic models applicable to evaluation of child abuse/neglect (CA/N) risk assessment models. (2) Assess the accuracy of the California Family Risk Assessment (CFRA) in predicting CA/N using the foregoing standards. (3) Compare the prediction accuracy of the CFRA with the prediction accuracy of coronary heart disease (CHD) prediction models.
Questions addressed: (1) What validation design and accuracy assessment standards are used to evaluate medical prognostic models? (2) What is the evidence for the accuracy of the CFRA using those standards? (3) How does the accuracy of the CFRA in predicting CA/N compare with the accuracy of CHD prediction models, which are a reasonable exemplar for the CA/N prediction effort?
Method: An external validation sample of 236 California reports of CA/N from San Luis Obispo and Sutter counties, and a larger temporal validation sample of 6307 California reports from Orange, Los Angeles, and Humboldt counties were investigated and assessed with the CFRA by line child welfare staff and were followed prospectively statewide for two years to discover reported, substantiated CA/N in any California county. CFRA accuracy in predicting substantiated CA/N was assessed by calibration and discrimination. Calibration was measured as the ratio of predicted to observed cases of CA/N seen during follow-up, with a ratio of 1.0 registering perfect calibration. Discrimination was measured by the area under the receiver operating characteristic (ROC) curve (AUC), with values from .60 to .85 found typical for medical prognostic models. CHD prediction literature was reviewed to acquire values of these accuracy measures for CHD prediction models. CFRA CA/N prediction accuracy and CHD prediction accuracy were then compared.
Results: Findings from external and temporal validation samples support the accuracy of CFRA prediction of CA/N. CFRA accuracy in predicting CA/N compared well with CHD prediction accuracy: (1) in the external validation sample, 43.42 CA/N cases were predicted during follow-up and 47 were observed, with consequent 7.6% deviation from perfect calibration. (2) In the temporal validation sample 857.49 CA/N cases were predicted and 801 were observed, with 7.1% deviation from perfect calibration. (3) The best performing of 20 Framingham CHD prediction models identified by systematic literature review predicted 222 CHD cases and 206 were observed, with 7.8% deviation from perfect calibration. (3) The CFRA external and temporal validation sample AUCs were .74 and .64, respectively. (4) For 26 CHD prediction cohorts found by literature review, the AUC mean and median values were .72 and .71, respectively, with a range from .60 to .84.
Conclusions/practice implications: (1) External and temporal validation results support the accuracy of the CFRA. (2) CFRA CA/N prediction accuracy parity with that for CHD prediction is encouraging, suggesting that wide use of the CFRA, properly implemented, could improve risk assessment accuracy in child protection. (3) Findings underline the importance of ensuring that no risk assessment model or method, including actuarial and consensus models and clinical judgment, is used in the field unless it has passed a test of external, or at least temporal validation.

1. Introduction

Every day, workers in child protective services (CPS) agencies across the United States examine reports of child abuse and neglect in order to decide which reports are serious enough to warrant use of the scarce resources available for intervention. (Kaufman, McIntire, & Santos, 2006, January 20; Ruttenberg & Mcintire, 2006, January 31). Knowing which children are most in need of help requires the ability to identify the children who will most likely be the victims of child abuse and neglect (CA/N) in the absence of

* Corresponding author.

http://dx.doi.org/10.1016/j.childyouth.2015.06.013
0190-7409/

effective intervention. Models that predict risk of CA/N with maximum possible accuracy are thus a requirement, not an option, for the rational practice of child protection.

In the past 30 years, considerable effort has been devoted to the development of actuarial risk assessment models designed to predict risk of CA/N. There is evidence of the validity and accuracy of these models (see e.g. Johnson & L'Esperance, 1984; Baird & Neunfeldt, 1988; Coohey, Johnson, Renner, & Easton, 2013). And there is evidence that they are more accurate than consensus-based models (Baird & Wagner, 2000) and unaided clinical judgment (Dawes, Faust, & Meehl, 1989; Grove & Meehl, 1996), which had been the dominant vehicle for risk assessment.

Now, half of American states (Akin, McDonald, & Tullis, 2010) and several other countries (Coohey et al., 2013) use CPS actuarial risk assessment models, apparently having perceived that they are more accurate.

Despite what appears to be progress, doubters remain. In the U.S., the other half of states has yet to adopt CPS actuarial risk assessment models, and a few continue to rely on clinical judgment alone in assessing risk of future CA/N (Akin et al., 2010). One prominent pair of critics (Jagannathan & Camasso, 2013) have argued that CPS actuarial risk assessment validation methodology and accuracy standards are insufficiently rigorous and, they imply, the field should adopt relevant standards from medicine (pp. 57, 58). Without distinguishing between types of risk assessment models (actuarial, consensus-based, or clinical judgment) these critics contend that "… CPS risk assessment has a long way to go to reach respectable levels of accuracy" (Jagannathan & Camasso, 2013, p. 58).

The inaction of half of American states regarding adoption of CPS actuarial risk assessment models despite the evidence that they are more accurate, and the published criticisms of CPS risk assessment models suggest that two important questions about CPS actuarial risk assessment remain and must be addressed, if it is to be more widely adopted.

First, how accurate are CPS actuarial risk assessment models compared with risk assessment models from other fields where "respectable" levels of risk assessment accuracy have been achieved?

Second, taking from the critics Jagannathan and Camasso the implied suggestion that medicine is a field that indeed has achieved respectable levels of accuracy that should be emulated, what are the medical risk assessment validation and accuracy standards that we should apply to CPS actuarial risk assessment models to ascertain whether these models have achieved "respectable" levels of accuracy.

Third, once we have measured the accuracy of CPS actuarial risk assessment models according to those validation and accuracy standards from medicine, how accurate are CPS actuarial risk assessment models compared with risk assessment models used in medicine, the implied exemplar of high standards of accuracy?

To begin to answer the questions above, we first identify through a narrative literature review some commonly used risk assessment validation design and accuracy assessment standards from medicine that are applicable to CA/N risk assessment models.

We next apply the above standards to one of the most commonly used CPS actuarial risk assessment models, the California Family Risk Assessment (CFRA), and compare its accuracy with the accuracy of coronary heart disease (CHD) prediction models from the Framingham Heart Study and elsewhere.

As part of a broader cardio-vascular risk assessment program, the American College of Cardiology (ACC)/American Heart Association (AHA) Task Force on Practice Guidelines in 2013 recommended that CHD risk assessment " … should be used … " for all American non-Hispanic whites and non-Hispanic African Americans 40 to 79 years of age, and " … may be considered … " for estimation of risk in American patients from populations other than "African Americans and non-Hispanic whites" (American Heart Association, 2014, p. S55).

2. Medical standards for the evaluation of risk assessment models, based on literature review

Risk assessment is "estimating … the probability of a future event or state. The outcome not only is unknown, but does not yet exist, distinguishing this task from diagnosis." (Cook, 2008, p. 17). In medicine, risk assessment is most often referred to as prognosis (Altman & Royston, 2000; Brindle, Beswick, Fahey, & Ebrahim, 2006; Rothwell, 2008). As is the case for CPS risk assessment models (see CFRA copy in Appendix A), the design of prognostic models generally involves using information (a set of potentially predictive variables) available during an assessment period for a sample of clients or patients, relating those variables to an observed outcome of interest, and creating a prognostic model (a collection of predictive variables) that statistically predicts the probability of experiencing an outcome of interest.

There are many ways to distill and combine information in order to create a tool (e.g., development of additive scales, factor analysis, maximum likelihood estimation), but these are always performed in the service of obtaining high levels of predictive validity. Thus, CPS actuarial risk assessment models are the conceptual, structural, and functional equivalent of medical prognostic models (Altman & Royston, 2000; Johnson, 2011; Rothwell, 2008; Wald & Woolverton, 1990).

To identify medical risk assessment validation and accuracy standards applicable to CPS actuarial risk assessment models, we reviewed the literature on the evaluation of medical prognostic model validity and accuracy.

2.1. Validation design standard for the evaluation of prognostic models

In an article entitled "What do me mean by validating a prognostic model", appearing in the journal *Statistics in Medicine* (Altman & Royston, 2000) we find that "Validity" refers to "generalizability", thus, "The idea of validating a prognostic model is generally taken to mean establishing that it works satisfactorily (i.e. with sufficient accuracy) for patients other than those from whose data it was derived." (Altman & Royston, 2000, p. 453).

Two types of design are recommended for validation of medical prognostic models. Though not the most rigorous, an acceptable validation design is *temporal validation*—use of a later sample taken from the place or places that contributed cases to a model's development sample (Altman & Royston, 2000), thus showing a model to be generalizable across time. The most rigorous validation design for medical prognostic models is *external validation*—use of a later sample from a new place or places, showing a model to be generalizable (accurate) across both time and with *differing case populations* (Altman & Royston, 2000; Altman, Vergouwe, Royston, & Moons, 2009; Brindle et al., 2006; Rothwell, 2008).

External and temporal validation of models require that the following steps be taken subsequent to creation of a model using a development sample. Necessary steps are specified by design type (Altman & Royston, 2000):

1. Temporal validation only: Validation is conducted using a new sample from the places (medical centers, government jurisdictions, etc.) that *contributed cases* to the model development sample.
2. External validation only: Validation is conducted using a new sample from a new place (medical center, government jurisdiction, etc.), that *contributed no cases* to the model development sample.
3. Both external and temporal validation: Predictions are made for cases in a new sample using the model's scoring system (additive scale, logistic regression equation, etc.) as built and specified using model development sample cases.
4. Both external and temporal validation: Predictions for cases in a new