ACRM
AMERICAN CONGRESS OF REHABILITATION MEDICINE

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2014;95:188–96

SPECIAL COMMUNICATION

# Statistical Analysis of Clinical Prediction Rules for Rehabilitation Interventions: Current State of the Literature

Anat Lubetzky-Vilnai, PT, MSc, Marcia Ciol, PhD, Sarah Westcott McCoy, PT, PhD, FAPTA

*From the Department of Rehabilitation Medicine, University of Washington, Seattle, WA*

**Abstract**

Deriving clinical prediction rules (CPRs) to identify specific characteristics of patients who would likely respond to certain interventions has become a research priority in physical rehabilitation. Understanding the appropriate statistical principles and methods of analyses underlying the derivation of CPRs is important for future rehabilitation research and clinical applications. In this article, we aimed to provide an overview of statistical techniques used for the derivation of CPRs to predict success following physical therapy interventions and to generate recommendations for improvements in CPR derivation research and statistical analysis in rehabilitation. We have summarized the current state of CPR intervention-related research by reviewing 26 studies. A common technique was found in most studies and included univariate association of factors with treatment success, stepwise logistic regression to determine the most parsimonious set of predictors for success, and calculation of accuracy statistics (focusing on positive likelihood ratios). We identified several shortcomings related to inadequate ratio of events by number of predictors, lack of standardization regarding acceptable interobserver reliability of predictors, questionable handling of predictors including reliance on univariate analysis and early categorization, and not accounting for dependence and collinearity of predictors in multivariable model construction. Interpretation of the derived CPRs was found to be difficult due to lack of precision of estimates and paradoxical findings when a subset of the predictors yielded a larger positive likelihood ratio than did the full set of predictors. Finally, we make recommendations regarding how to strengthen the use of statistical principles and methods to create consistency across rehabilitation research for CPR derivations.

Archives of Physical Medicine and Rehabilitation 2014;95:188–96



Clinical prediction rules (CPRs) are clinical tools that quantify the individual contribution of medical history, physical examination, and basic laboratory results toward diagnosis, prognosis, or response to treatment of an individual patient.[1] In the past 5 years, systematic reviews, original studies, and letters to editors[2] indicate increased interest in CPRs in physical rehabilitation research.[3] For example, several clinical trials investigating interventions performed by physical therapists have not been conclusive, making the interventions' adoption in clinical settings difficult to support.[3,4] The authors have suggested that these studies included heterogeneous groups of patients, with mixed response to the intervention. Therefore, identifying characteristics of patients who would likely respond to certain interventions has become a research priority.[5]

The first step in developing a CPR is derivation,[6] and it includes 2 phases: hypothesis setting and hypothesis testing.[6] Hypothesis setting phase studies aim to identify combinations of factors that predict a diagnosis, prognosis, or outcome following an intervention. The quality of hypothesis setting derivation studies should be considered before CPRs are further researched and implemented into clinical practice.[7] If the research methods used to derive a rule are not robust, the resultant rule might be biased or misleading[8] and little weight can be given to its conclusions.[9] The derivation of CPRs is mostly based on statistical techniques and, therefore, understanding the statistical principles and analysis underlying the derivation of CPRs is important for both clinicians and researchers.

Recent reviews have discussed quality issues related to the development of physical rehabilitation CPRs and highlighted how methodological aspects can affect the interpretation and clinical application of these rules. The topics discussed included hierarchy and stages of CPR development,[1-3,10] optimal study design for CPR derivation,[1-3,10-12] and sample size.[3,7,10] Statistical analysis

No commercial party having a direct financial interest in the results of the research supporting this article has conferred or will confer a benefit on the authors or on any organization with which the authors are associated.

0003-9993/14/$36 - see front matter © 2014 by the American Congress of Rehabilitation Medicine
http://dx.doi.org/10.1016/j.apmr.2013.08.242

used for CPR derivation and issues related to model construction and evaluation have not been appraised.

In light of these findings, the aims of this article were to provide (1) an overview of statistical techniques that have been used for the derivation of CPRs in physical rehabilitation, using examples from the literature on physical therapy interventions; and (2) recommendations for improvements in CPR derivation methodology to predict success with rehabilitation interventions.

## Methods

### Search strategy

The 3 authors independently performed a systematic search in PubMed, CINAHL, and Web of Science. The following combinations of key words were used: ("clinical prediction rule" OR "clinical prediction rules" OR "clinical decision rule" OR "clinical decision rules") AND physical therapy. Studies identifying characteristics of patients likely to benefit from interventions performed by physical therapists, published in English through November 2012, were included in this article. Reference lists of all articles and reviews were also manually searched for additional titles.

## Results

Appendix 1 shows the article selection process with exclusions for the following reasons: not intervention related; assessing response to interventions not routinely performed by physical therapists (eg, chiropractic medicine); validation study; reviews; and letters to the editors/commentaries. The 26 publications included studies on low back pain (7 studies),[13-19] neck pain (5 studies),[20-24] headaches (3 studies),[25-27] hip osteoarthritis (1 study),[28] patellofemoral pain (4 studies),[29-32] knee osteoarthritis (2 studies),[33,34] tennis elbow (1 study),[35] carpal tunnel syndrome (1 study),[36] ankle sprain (1 study),[37] and work-related pain (1 study).[38]

In the following sections, we critically discuss various aspects of the derivation of CPRs. Definitions of statistical terms used in this article are given in appendix 2.

### Study design

Twenty studies used a prospective cohort, single-group design. The remaining 6 studies used data from randomized controlled trials but analyzed them as a single-group study and did not test for treatment effect modifiers.[15,25,28,32,34,35] In 2 studies, diagnostic accuracy statistics of the CPR were also calculated in the control group to confirm that their rule predicted response to the intervention rather than natural history of the disorder.[28,35] For a detailed discussion of optimal study design at the derivation stage of intervention-related CPRs, see Hancock et al.[8]

***List of abbreviations:***

| | |
|---|---|
| **CI** | **confidence interval** |
| **CPR** | **clinical prediction rule** |
| **LR+** | **positive likelihood ratio** |
| **LR−** | **negative likelihood ratio** |
| **VIF** | **variance inflation factor** |

### Sample size

The most common rule of thumb used in the literature for determining sample size in regression analysis is to use 10 or 15 observations for each predictor in the model.[39] Of the 26 studies reviewed, only 3 did not have at least 10 participants per final predictor.[26,27,31] However, the initial "10:1" rule was based on a simulation study performed by Peduzzi et al[40] designed to evaluate the number of events (death, in Peduzzi's study[40]) needed per explanatory variable in a logistic regression analysis. They found that the regression coefficients were biased in both positive and negative directions and had large variance when having fewer than 10 events per predictor. The final sample size depended not only on the 10 events per variable ratio but also on the probability of the event over the study period. For example, if one had 5 explanatory variables and an event's probability (success) was 0.4, then one would need $5\times10=50$ events to occur, and therefore, 125 ($=50/0.4$) observations would be needed. In addition, this rule is concerned with stability of the estimates and not with statistical power. A thorough discussion of this rule appears in van Belle.[41]

In the studies reviewed, the consistent use of this rule of thumb indicates that the current CPR literature fails to acknowledge the complexity of the sample size issue[8,39] and most likely have used fewer participants than needed for CPR derivation. Sample size required in a study depends on the type of outcome (continuous, discrete, ordinal, nominal), its variation in the population of interest, the research question, the statistical analysis used, the effect size, and the statistical power desired to detect the effect.[39] While hypothesis-setting derivation studies are exploratory in nature, a researcher should aspire for samples sufficiently large to contain enough variation to cover all possible combinations of predictors in the model. When samples are divided into subgroups and some subgroups have no participants, it creates numerical modeling problems that will likely be reflected in coefficients that have unreasonably large standard errors.[39] Below, we discuss some shortcomings (eg, wide confidence intervals [CIs] around the point estimates and difficulty in the clinical interpretation of the CPR) that might result from using inadequate sample sizes.

### Reliability of predictors

CPRs should be based on predictors that are valid (measure what one wants to measure), reliable (can be repeated consistently), and diagnostically meaningful.[42] The studies reviewed attempted to select predictors with literature evidence of being reliable and valid. In addition, 9 studies conducted their own interobserver reliability assessments for a few physical examination items,[13,21,28-33,37] but only 1 study excluded predictors with poor reliability.[28] For example, Lesher et al[30] had 2 predictors in their final model: one with an intraclass correlation of .66 and another with a kappa coefficient of only .49. Sutlive et al[29] stated that the generally low interobserver reliability values found for their measurements posed a threat to the internal validity of their study and might limit the application of the CPR they developed.

Lack of interobserver reliability assessment seems to also be a problem in CPRs from other medical fields. For example, Serrano et al[43] systematically reviewed the quality of clinical decision rules for syncope in the emergency department. They reported that the lack of interobserver reliability assessment was the most common methodological weakness in the 18 studies reviewed, with only 2 derivation studies assessing reliability.