



City Research Online

City, University of London Institutional Repository

Citation: Ban, J-W., Emparanza, J. I., Urreta, I. & Burls, A. (2016). Design Characteristics Influence Performance of Clinical Prediction Rules in Validation: A Meta-Epidemiological Study. PLoS One, 11(1), e0145779. doi: 10.1371/journal.pone.0145779

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/13167/>

Link to published version: <https://doi.org/10.1371/journal.pone.0145779>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

RESEARCH ARTICLE

Design Characteristics Influence Performance of Clinical Prediction Rules in Validation: A Meta-Epidemiological Study

Jong-Wook Ban^{1*}, José Ignacio Emparanza², Iratxe Urreta², Amanda Burls³

1 Evidence-Based Health Care Programme, Department of Continuing Education, Kellogg College, University of Oxford, Oxford, United Kingdom, **2** CASPe, CIBER-ESP, Clinical Epidemiology Unit, Hospital Universitario Donostia, San Sebastian, Spain, **3** School of Health Sciences, City University London, London, United Kingdom

* jong-wook.ban@conted.ox.ac.uk



Abstract

Background

Many new clinical prediction rules are derived and validated. But the design and reporting quality of clinical prediction research has been less than optimal. We aimed to assess whether design characteristics of validation studies were associated with the overestimation of clinical prediction rules' performance. We also aimed to evaluate whether validation studies clearly reported important methodological characteristics.

Methods

Electronic databases were searched for systematic reviews of clinical prediction rule studies published between 2006 and 2010. Data were extracted from the eligible validation studies included in the systematic reviews. A meta-analytic meta-epidemiological approach was used to assess the influence of design characteristics on predictive performance. From each validation study, it was assessed whether 7 design and 7 reporting characteristics were properly described.

Results

A total of 287 validation studies of clinical prediction rule were collected from 15 systematic reviews (31 meta-analyses). Validation studies using case-control design produced a summary diagnostic odds ratio (DOR) 2.2 times (95% CI: 1.2–4.3) larger than validation studies using cohort design and unclear design. When differential verification was used, the summary DOR was overestimated by twofold (95% CI: 1.2–3.1) compared to complete, partial and unclear verification. The summary RDOR of validation studies with inadequate sample size was 1.9 (95% CI: 1.2–3.1) compared to studies with adequate sample size. Study site, reliability, and clinical prediction rule was adequately described in 10.1%, 9.4%, and 7.0% of validation studies respectively.

OPEN ACCESS

Citation: Ban J-W, Emparanza JI, Urreta I, Burls A (2016) Design Characteristics Influence Performance of Clinical Prediction Rules in Validation: A Meta-Epidemiological Study. PLoS ONE 11(1): e0145779. doi:10.1371/journal.pone.0145779

Editor: Ewout W Steyerberg, University Medical Center Rotterdam, NETHERLANDS

Received: June 20, 2015

Accepted: December 8, 2015

Published: January 5, 2016

Copyright: © 2016 Ban et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: These authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Conclusion

Validation studies with design shortcomings may overestimate the performance of clinical prediction rules. The quality of reporting among studies validating clinical prediction rules needs to be improved.

Introduction

Clinical prediction rules help clinicians address uncertainties surrounding the diagnosis, prognosis or response to treatment using information from individual patient's history, physical examination and test results [1–3]. Contrary to the traditional approach where intuition is typically used to handle clinical uncertainties, clinical prediction rules enable clinicians to explicitly integrate information from individual patients and estimate the probability of an outcome.

Once a clinical prediction rule is constructed in a derivation study by combining variables predictive of an outcome, the reproducibility and generalizability of the clinical prediction rule should be evaluated in validation studies [4–8]. A clinical prediction rule that performed well in a derivation may not fare so well when it is applied to different populations or settings [6, 9–11]. Therefore, only clinical prediction rules that have proven performance through external validations should be trusted and considered for application in clinical practice [5, 12].

There have been several methodological standards proposed over the past three decades that considered design and reporting characteristics of studies deriving, validating as well as assessing the impact of clinical prediction rules [1–3, 5, 8, 12, 13]. Despite the available methodological standards, overall methodological qualities of clinical prediction rule research described in previous reports have been far from optimal [1, 2, 14–17]. However, the findings of these reports were largely based on the evaluation of derivation studies while including a limited number of validation studies. Recently, a systematic review of multivariable prediction models collected from core clinical journals showed that important methodological characteristics are poorly described in validation studies [18].

There is a growing body of empirical evidence showing that the design and conduct of a study can influence the results. For example, a number of meta-epidemiological studies which examined clinical trials included in meta-analyses have shown that failure to ensure proper random sequence generation, allocation concealment or blinding can lead to the overestimation of treatment effects [19–22]. In diagnostic test accuracy studies, it has been suggested that the use of less than optimal study design characteristics such as retrospective data collection, nonconsecutive subject selection or case-control design may lead to overestimated test accuracy [23, 24]. For validation studies of clinical prediction rules, the potential implications of employing design characteristics that are not compatible with currently available methodological standards are yet to be determined.

Our primary objective was to evaluate whether validation studies conducted using design characteristics that are inconsistent with methodological standards are associated with the overestimation of predictive performance. We also aimed to estimate the proportion of published validation studies that clearly reported important methodological characteristics so that the readers could assess the validity.

Materials and Methods

Reporting and design characteristics of studies validating clinical prediction rule

The methodological standards for clinical prediction rules [1–3, 13] as well as quality assessment tools and a reporting guideline for diagnostic test accuracy studies [25–27] were reviewed

Table 1. Definitions of (a) reporting and (b) design characteristics.

(a) Reporting characteristic		Definition
Population	Age, sex and important clinical characteristics are described	
Study site	Geographic location, institution type and setting, and how patients were referred are described	
Prediction rule	Clear and detailed descriptions of predictor variables and prediction rule are provided with a description of process to ensure accurate assessment of rule such as training	
Reliability	Intra-observer or inter-observer reliability of prediction rule is described	
Outcome	Clear and detailed definition of outcome is provided	
Results	Estimates of predictive performance are presented with confidence intervals	
Follow up	Clearly describes what happened to all enrolled patients	
(b) Design characteristic	Level	Definition
Sample size	Adequate	At least 100 patients with outcome and 100 patients without outcome
	Inadequate	Less than 100 patients with outcome or 100 patients without outcome
Patient selection	Consecutive	All consecutive patients meeting inclusion criteria are selected
	Nonconsecutive	Selection methods other than consecutive selection were used
Disease spectrum	Cohort	Prediction is ascertained before outcome is determined
	Case-control	Prediction is ascertained after outcome is determined
Validation type	Broad	Validation in different settings, with different patients and by different clinicians
	Narrow	Validation in similar settings, with similar patients or by similar clinicians
	Internal	Validation using methods such as split-sample validation, cross-validation or bootstrap
Assessment	Blind	Assessment of prediction without the knowledge of outcome and assessment of outcome without the knowledge of prediction
	Non-blind	Assessment of prediction with the knowledge of outcome or assessment of outcome with the knowledge of prediction
Verification	Complete	All predictions are verified using the same reference standard
	Partial	Only a subset of predictions are verified
	Differential	A subset of predictions are verified using different reference standard
Data collection	Prospective	Data collection for validation study is planned before prediction and outcome are assessed
	Retrospective	Prediction and outcome are assessed before data collection for validation starts

doi:10.1371/journal.pone.0145779.t001

to identify reporting and design characteristics of studies validating clinical prediction rules. Definitions of 7 reporting characteristics and 7 design characteristics examined in our study are outlined in [Table 1](#).

Simulations have shown that validation studies with less than 100 patients with and without an outcome may not identify the invalidity of a regression model [11, 28]. Case-control design, nonconsecutive enrollment, and retrospective data collection may lead to a biased selection of patients [29, 30]. Case-control design was associated with the overestimation of diagnostic test accuracy in a meta-epidemiological study [23]. Case-control design may be obvious when patients with clinical suspicion and healthy subjects without clinical suspicion are recruited separately [31]. However, it may be indistinct when an outcome is determined before the prediction is assessed in a “reversed-flow” design [31]. Knowing the result of the prediction may influence the assessment of an outcome and knowing the outcome of a patient may influence estimating the prediction [1–3, 13] although blinding did not significantly influence diagnostic test accuracy in meta-epidemiological studies [23, 24]. Partial or differential verification of predictions may lead to verification bias. Differential verification was associated with the overestimation of diagnostic accuracy [23, 24]. Even though validation type is not a methodological standard, this feature was included in our meta-epidemiological analysis since the type of validation is likely to influence the performance of a clinical prediction rule in validation.

Selection of systematic reviews and validation studies

Medline, EMBASE, the Cochrane library and the Medion database (www.mediondatabase.nl) were searched for systematic reviews of clinical prediction rule studies. Searches were limited to the systematic reviews published between 2006 and 2010. No language restriction was applied. All electronic data base searches were conducted on 11 November 2011.

The search strategies for electronic databases were constructed by combining terms from validated filters for clinical prediction rule study and systematic review [32–35]. Search strategies for Medline (OvidSP™, 1946 to 2011), EMBASE (OvidSP™, 1974 to 2011), the Cochrane library, and the Medion database (www.mediondatabase.nl) are presented in S1 Fig.

Selection of studies was carried out in two steps. First, eligible systematic reviews of clinical prediction rule studies were identified from the results of the electronic database search. Then, eligible validation studies of clinical prediction rule included in these systematic reviews were selected for data extraction.

A systematic review of clinical prediction rule studies was eligible if:

1. it examined a clinical prediction rule which is defined for this study as a tool that produces a probability of an outcome to help clinicians with the diagnosis, prognosis or treatment decision by combining three or more predictor variables from patient's history, physical examination or test results,
2. it meta-analyzed the results of clinical prediction rule studies,
3. at least one meta-analysis pooled the results of four or more validation studies of a same clinical prediction rule,
4. a diagnostic 2 by 2 table could be constructed from four or more validation studies, and
5. the primary studies were not selected or excluded based on any of the seven study design characteristics examined in this study.

Once a systematic review of clinical prediction rule studies was determined eligible, full text articles of all clinical prediction rule studies included in the systematic review were reviewed. A clinical prediction rule study was included if:

1. it was a validation study of clinical prediction rule in which the generalizability of an existing clinical prediction rule was assessed by comparing its prediction with a pre-specified outcome, and
2. the information from the study allowed for the construction of a diagnostic 2 by 2 table.

Data extraction and analysis

From each eligible validation study, a diagnostic 2 by 2 table was constructed and the reporting and design characteristics were recorded. A systematic review may contain multiple meta-analyses that pooled the results of clinical prediction rule studies using various thresholds. Therefore, the following strategy was used to identify an optimal threshold for data extraction from a validation study. The strategy was aimed at including the maximum number of validation studies while minimizing heterogeneity of data due to threshold effects and preventing a validation study from being included in multiple analyses.

1. when the results of clinical prediction rule studies were meta-analyzed at a single threshold, this threshold was used for our data analysis.
2. when the results of clinical prediction rule studies were meta-analyzed at multiple common thresholds, the threshold that would maximize the number of eligible validation studies was chosen for data extraction.

3. when the results of clinical prediction rule studies with various thresholds were meta-analyzed and the thresholds of validation studies used in meta-analysis are known, each threshold that was chosen by authors of the systematic review to conduct meta-analysis was used.

Meta-epidemiological studies evaluate biases related to study characteristics by collecting and analyzing many meta-analyses. A multilevel logistic regression is commonly used in meta-epidemiological studies but it may underestimate standard errors and is mathematically complex since an indicator variable is required for each primary study and each meta-analysis [36]. We used a “meta-meta-analytic approach” to examine the influence of design characteristics on the performance of clinical prediction rule which involves two steps, a meta-regression and a meta-analysis of regression coefficients [36].

Firstly, a multivariable meta-regression was conducted in each meta-analysis for seven design characteristics using Meta-Disc [37] which carries out a meta-regression by adding covariates to the Moses-Littenberg model [38, 39]. The output of the meta-regression in Meta-Disc includes a coefficient which is the natural logarithm of relative diagnostic odds ratio (RDOR) and a standard error of coefficient. Performance of a clinical prediction rule in a validation study can be presented with a diagnostic odds ratio (DOR). A relative diagnostic odds ratio (RDOR) is a ratio of the DOR of a prediction rule with a certain study characteristic (e.g. case-control design) and the DOR of a prediction rule without the study characteristic. In other words, a RDOR represents the influence of a study characteristic on the performance of clinical prediction rules. As a significant amount of heterogeneity was expected between validation studies of a clinical prediction rule, meta-regressions were conducted under the random-effects assumption.

Secondly, the results of meta-regressions were meta-analyzed using Stata 12 [40]. The regression coefficients computed in the first step were meta-analyzed using the DerSimonian and Laird option in Metaan command [41]. Then, the summary RDORs were obtained from the antilogarithms of the meta-analyses results. A significant heterogeneity was assumed to be present between the results of meta-regressions and random-effects meta-analyses were conducted.

The search, selection, data extraction were completed by one of the authors (J.B.). Two clinical epidemiologists (J.E. and I.U.) independently extracted a second set of data from a random sample of validation studies included in data analysis for comparison. The agreements on the assessments of six design characteristics were evaluated using Cohen’s kappa.

Results

Selection of material and description of sample

The selection processes for eligible systematic reviews and validation studies are illustrated in [S2 Fig](#). From 46848 references found in the electronic database search, 15 eligible systematic reviews consisting of 31 meta-analyses were identified [42–56]. There were 287 validation studies from 229 clinical prediction rule articles included in these meta-analyses that met the inclusion criteria.

The description of the 15 systematic reviews included in the study are outlined in [S1 Table](#). Although the systematic reviews of clinical prediction rule published between 2006 and 2010 were searched, 13 of 15 systematic reviews in the sample were from either year 2009 or 2010. Eleven systematic reviews evaluated diagnostic prediction rules whereas four systematic reviews examined prognostic prediction rules. Meta-analyses generally contained a small number of validation studies. Twenty one meta-analyses (67.7%) included 10 or less validation studies. Only one (3.2%) meta-analysis was conducted with more than twenty validation

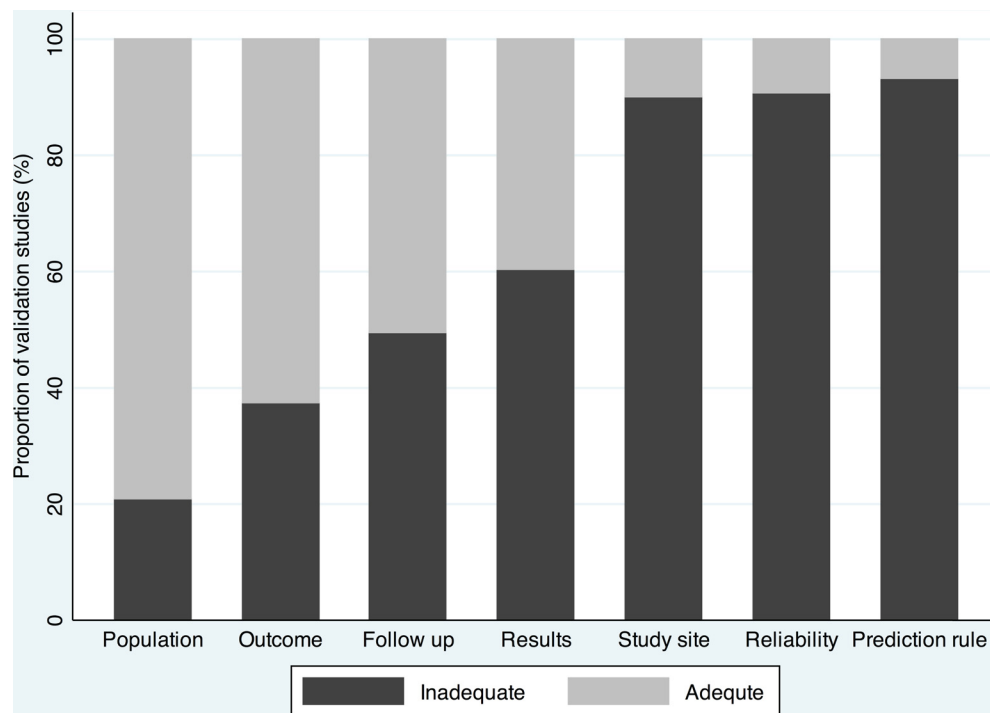


Fig 1. Quality of reporting. Proportion of validation studies with adequate and inadequate description of reporting characteristics.

doi:10.1371/journal.pone.0145779.g001

studies. Fifteen meta-analyses evaluated objective outcomes such as the presence of ovarian malignancy [46] whereas sixteen meta-analyses evaluated subjective outcomes such as the diagnosis of postpartum depression [47].

Two hundred and seventy eight validation studies (96.9%) were written in English. There were only 12 validation studies conducted before 1991. The number of validation studies increased to 110 between 1991 and 2000 and to 165 between 2001 and 2010.

Quality of reporting in validation studies of clinical prediction rule

As presented in Fig 1, validation studies providing insufficient descriptions of reporting characteristics were common. Study site, reliability, and clinical prediction rule were inadequately reported in the majority of validation studies. For reporting of study site, 118 (41.1%), 113 (39.3%) and 241 (84.0%) validation studies provided insufficient descriptions of location, institution type and referral to study site respectively. For reporting of clinical prediction rule, the processes to ensure an accurate application of prediction rule such as training were described in 20 (7.0%) validation studies. Predictive performance was reported with confidence interval in 84 (36.7%) validation studies. Although 145 (50.5%) validation studies provided information on the follow up of patients, flow diagram was used only in 23 (8.0%) validation studies.

There were 66 (23%) validation studies meeting 0 or 1 reporting characteristic and 147 (51.2%) validation studies satisfying 2 or 3 reporting characteristics. Although only 74 (25.8%) validation studies met 4 or more reporting characteristics, the proportion of validation studies satisfying 4 or more reporting characteristics increased from 8.9% between 1991 and 1995, to 27.7% between 1996 and 2000, to 25.9% between 2001 and 2005, and to 42.1% between 2006 and 2010 as seen in Fig 2(A).

Influence of design characteristics on predictive performance of clinical prediction rule

Design shortcomings and insufficient descriptions of design characteristics were prevalent among validation studies of clinical prediction rule as presented in [S2 Table](#).

There were 53 (18.5%) validation studies meeting 0 or 1 design characteristic and 161 (56.1%) validation studies satisfying 2 or 3 design characteristics. There were 73 (25.4%) validation studies conducted while satisfying 4 or more design characteristics. The proportion of validation studies satisfying 4 or more reporting characteristics did not improve over time as seen in [Fig 2\(B\)](#). Between 2006 and 2010, there were 15 (26.3%) validation studies that satisfied 4 or more design characteristics.

Fisher's exact test showed that patient selection ($p = 0.028$), disease spectrum ($p = 0.008$), and data collection ($p = 0.00$) significantly varied between validation studies of diagnostic and prognostic prediction rules. For example, nonconsecutive selection was more common among diagnostic rules (17.5%) compared to prognostic rules (4.8%). Case-control design was also more prevalent among diagnostic rules (21.6%) compared to prognostic rules (4.8%). On the other hand, retrospective data collection was more frequently found in prognostic rules (50%) than diagnostic rules (16.7%). Although there was no significant overall variation in verification between diagnostic and prognostic rules, all studies using differential verification validated diagnostic prediction rules.

Validation studies using inadequate sample size and differential verification were associated with the overestimation of predictive performance in univariable analysis as presented in [S2 Table](#). In order to construct a multivariable model, some of the levels for design characteristics were collapsed according to the summary RDORs from the univariable analysis. Consecutive selection and unclear patient selection were collapsed and compared with non-consecutive selection. Case-control design was compared after cohort design and unclear spectrum were combined. Narrow validation was compared with a combined category of broad validation and unclear validation type. Differential verification was compared with a combined category of complete, partial, and unclear verification. Lastly, prospective and retrospective data collection were collapsed and compared with unclear data collection according to the summary RDORs from the univariable model. Blind assessment of prediction and outcome was not clearly

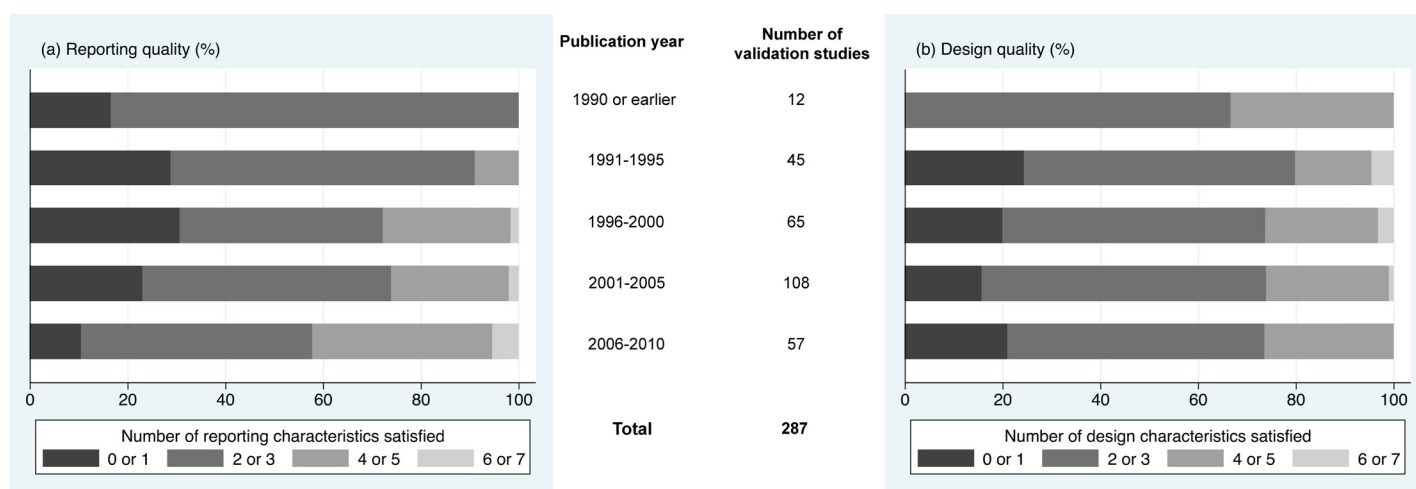


Fig 2. Evolution of methodological quality over time. Proportion of validation studies satisfying (a) reporting characteristics and (b) design characteristics recommended in methodological standards.

doi:10.1371/journal.pone.0145779.g002

reported in 207 (72.1%) of the validation studies. Therefore, this feature was excluded in the multivariable model.

The results of the multivariable analysis are presented in Fig 3. Validation studies using case-control design produced the largest summary RDOR of 2.2 (95% CI: 1.2–4.3) compared to the validation studies with cohort design and unclear disease spectrum when controlled for the influence of other design characteristics in the multivariable model. The summary RDOR for differential verification compared to complete, partial, and unclear verification was 2.0 (95% CI: 1.2–3.1) indicating that validation studies using differential verification are likely to overestimate DOR by twofold. Validation studies conducted with inadequate sample size produced a summary RDOR of 1.9 (95% CI: 1.2–3.1).

The summary RDOR of narrow validation was 1.8 with a large 95% confidence interval of 0.8 to 4.4. When an outlier meta-analysis that produced a large coefficient in meta-regression was excluded, the summary RDOR of narrow validation was reduced to 1.3 with a less wide 95% confidence interval of 0.6 to 2.8. The outlier meta-analysis summarized performance of HADS, the Hospital Anxiety and Depression Scale, in palliative care setting [55].

Validation studies with unclear data collection methods produced a summary RDOR that was 20% larger than the summary RDOR of validation studies that clearly used prospective or retrospective data collection. However, the 95% confidence interval of the summary RDOR for validation studies with unclear data collection was 0.3 to 5.0. The wide 95% confidence interval may have been caused by the small number of heterogeneous meta-analyses ($I^2 = 65.1\%$) contributing to the estimation of the summary RDOR in this category.

Random-effects assumption between the results of meta-regressions was tested in sensitivity analysis. The conclusions in univariable and multivariable models were unaffected under fixed effect assumption. From 78 randomly selected validation studies, a second set of data were extracted independently for comparison. Cohen's kappa was 0.64, 0.39, 0.05, 0.24, 0.00 and 0.26 for sample size, patient selection, spectrum, validation type, verification and data collection.

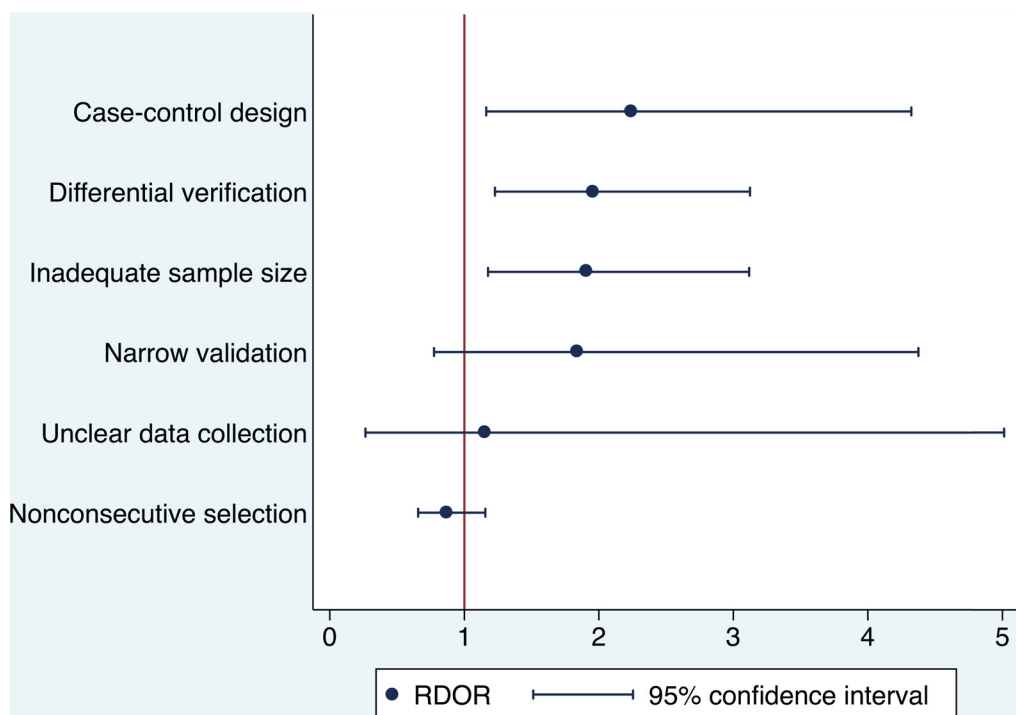


Fig 3. Influence of design characteristics on the performance of clinical prediction rule in multivariable analysis.

doi:10.1371/journal.pone.0145779.g003

Discussion

In this meta-epidemiological study, we investigated whether validation studies with less than optimal design characteristics are associated with the overestimation of predictive performance. We also evaluated whether studies validating clinical prediction rules are published while providing sufficient descriptions of design and reporting characteristics recommended in methodological standards.

Our results showed that many validation studies are conducted using design characteristics inconsistent with recommendations from methodological standards. The results also demonstrated that validation studies with design shortcomings overestimate the predictive performance. Among the design characteristics examined, case-control design was associated with the largest overestimation of predictive performance. This is consistent with the previous observation by Lijmer et al. [23] that using case-control design in diagnostic test accuracy studies lead to the largest overestimation of test accuracy. Case-control design was uncommon in diagnostic test accuracy studies according to the previous reports by Lijmer et al. [23] and Rutjes et al. [24] where it was used in 2.3% and 8.6% of their studies respectively. It is concerning since case-control design was used in 19.2% of validation studies of clinical prediction rule in our study. We used the definition of case-control design by Rutjes et al. [31] which included a "reversed-flow" design where an outcome is determined before the prediction is assessed. The use of this definition may have lead to the detection of validation studies using less obvious case-control design. Furthermore, the higher prevalence of case-control design in our sample may reflect lower overall methodological quality of validation studies compared to diagnostic test accuracy studies. Readers should pay close attention to detect case-control design which is relatively common among validation studies of clinical prediction rule and should use caution interpreting the results if a clinical prediction rule was validated using case-control design.

Although differential verification was less commonly observed compared to case-control design in the present study sample, it was also associated with substantial overestimation of predictive performance. This finding is similar to the previous observation by Lijmer et al. [23] in diagnostic test accuracy studies. Clinicians evaluating the performance of clinical prediction rule should be careful with the results of a validation study when subsets of predictions were verified with different reference standards.

The results showed that 87.1% of validation studies used less than sufficient number of subjects for validation of logistic regression model. These studies with inadequate sample size were associated with the overestimation of predictive performance by almost twofold. This is a confirmation of findings from simulations by Vergouwe et al. [28] and Steyerberg [11] that at least 100 subjects with the outcome and 100 subjects without the outcome were needed to detect a modest invalidity in logistic regression models.

Although it is logical to presume that clinical prediction rules would perform less well in broad validation, our study failed to observe an association between the validation type and the performance of clinical prediction rule when controlled for the influence of other design characteristics. It was common to find that all or nearly all studies included in a meta-analysis were classified as the same validation type (e.g. narrow validation) and a regression coefficient could not be generated for this feature in many meta-regressions. Meta-analyzing a small number of regression coefficients may have resulted in the imprecise estimation of a summary RDOR. Nonconsecutive selection does not appear to influence the performance of clinical prediction rule in validation studies. This finding is similar to previous observations by Lijmer et al. [23] and Stengel et al. [57].

Potential clinical implication of findings from the present study is discussed in the following example. According to one of the articles in the present study sample that validated

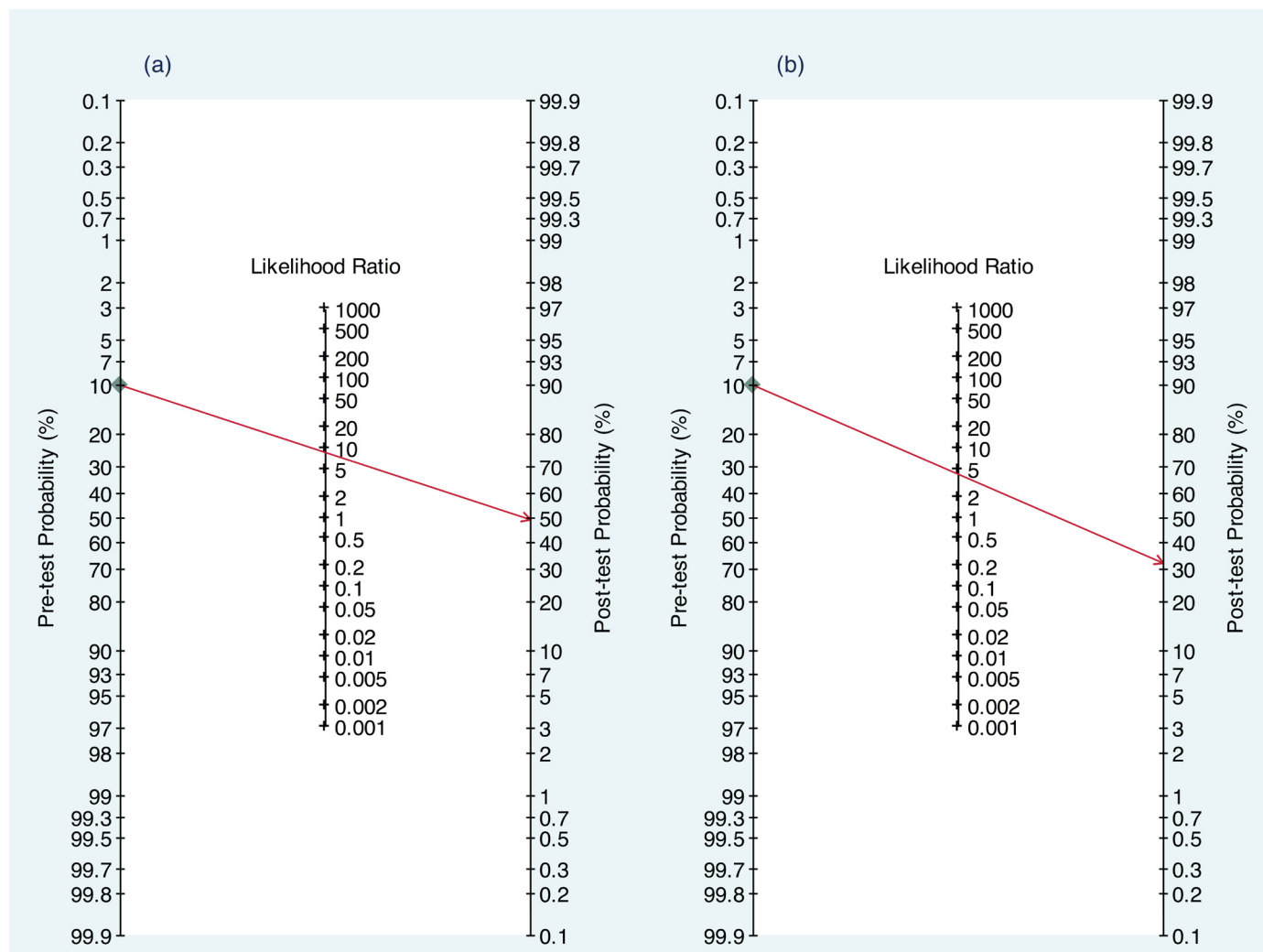


Fig 4. Fagan nomogram. Applying the sensitivity and specificity of (a) 90% as presented in the validation study [58] and (b) 81% from an unbiased study to a patient with 10% probability of rheumatoid arthritis.

doi:10.1371/journal.pone.0145779.g004

classification criteria for rheumatoid arthritis in an outpatient setting, the sensitivity and specificity of 1987 American College of Rheumatology (ACR) criteria were both 90% [58]. However, the DOR of 81 calculated from the sensitivity and specificity may have been overestimated by 4.2 fold since this study was conducted using inadequate sample size as well as case-control design. This means that if adequate sample size and cohort design were used, the DOR would have been about 19 with corresponding sensitivity and specificity of 81%.

Fig 4 illustrates potential clinical consequences of applying biased results using a hypothetical patient with 10% risk of rheumatoid arthritis. If the presence of rheumatoid arthritis is predicted with 1987 ACR criteria using the sensitivity and specificity of 90% as presented by the authors [58], the probability of rheumatoid arthritis in this patient would be 50%. On the other hand, applying the sensitivity and specificity of 81% from a hypothetical unbiased study, the probability of rheumatoid arthritis in this patient would be only 32%. In summary, applying the potentially biased results from validation studies with design shortcomings may produce overly confident clinical predictions. Consequently, unnecessary or potentially harmful therapy may be given to patients and life saving intervention may be withheld.

The reporting quality of validation studies included in the present study sample was poor. Inadequate descriptions of study site, reliability, and clinical prediction rule were highly prevalent. The results presented with confidence intervals and follow-up of study participants using flow diagram were found only in a small number of validation studies which is consistent with the findings from a recent systematic review [18]. Many validation studies did not describe whether blind assessments of clinical prediction rule and outcome were carried out.

For the findings of validation studies to be trusted and to be safely applied to clinical practice, the design, conduct and analysis of validation study must be clearly and accurately described. Unfortunately, the existing evidence including our study indicates key information is often missing or not clearly reported in validation studies. Collins et al. [59] published the TRIPOD (Transparent Reporting of multivariable prediction model for Individual Prognosis Or Diagnosis) statement which is a 22-item checklist for reporting derivation and validation of clinical prediction models. Although many methodological features evaluated in our study are included in the TRIPOD statement, there are items that are not examined in our study. Future research should examine the impact of TRIPOD statement in reporting of prediction rule studies.

The main limitation of this study is that only data from a limited number of validation studies could be independently assessed by other reviewers. Restraints in time and funding prohibited complete independent verification of data. Furthermore, there were considerable disagreements in interpretations of design characteristics between reviewers. Unfortunately, clear resolution of the disagreement was impossible for most cases due to ambiguity in description of these design characteristics in validation studies. Secondly, the small number of validation studies included in meta-analyses of clinical prediction rule studies prohibited incorporating more covariates in the multivariable model. As meta-analyses of clinical prediction rule studies become more readily available, sampling the meta-analyses with a greater number of validation studies may allow for the examination of additional covariates that influence the predictive performance. Finally, a meta-meta-analytic approach used in our study does not formally account for the clustering of DORs among clinical domains or clinical prediction rules compared to a multilevel regression modeling approach. Although this method performed well in a meta-epidemiological study of clinical trials [36], its robustness is not known in handling heterogeneous validation studies of prediction rules and meta-analyses of validation studies.

In conclusion, our study demonstrated that many validation studies of clinical prediction rule are published without complying with reporting and design characteristics recommended in the methodological standards. The results of validation studies should be interpreted with caution if design shortcomings such as use of case-control design, differential verification or inadequate sample size are present as the predictive performance may have been overestimated. Conscious efforts are needed to improve the quality of design, conduct and reporting of validation studies.

Supporting Information

S1 Fig. Search strategies for systematic reviews of clinical prediction rule studies.
(PDF)

S2 Fig. Flow diagram. Selection of (a) systematic reviews and (b) validation studies.
(PDF)

S1 File. Dataset for multivariable model.
(XLSX)

S1 Table. Systematic reviews of clinical prediction rule studies included in analysis.
(PDF)

S2 Table. Univariable analysis. Influence of design characteristics on the performance of clinical prediction rule.
(PDF)

Author Contributions

Conceived and designed the experiments: JB JE IU AB. Performed the experiments: JB JE IU AB. Analyzed the data: JB JE IU AB. Contributed reagents/materials/analysis tools: JB JE IU AB. Wrote the paper: JB JE IU AB.

References

1. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. *The New England journal of medicine*. 1985; 313(13):793–9. Epub 1985/09/26. doi: [10.1056/NEJM198509263131306](https://doi.org/10.1056/NEJM198509263131306) PMID: [3897864](https://pubmed.ncbi.nlm.nih.gov/3897864/).
2. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA: the journal of the American Medical Association*. 1997; 277(6):488–94. Epub 1997/02/12. PMID: [9020274](https://pubmed.ncbi.nlm.nih.gov/9020274/).
3. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA: the journal of the American Medical Association*. 2000; 284(1):79–84. Epub 2000/06/29. PMID: [10872017](https://pubmed.ncbi.nlm.nih.gov/10872017/).
4. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Annals of internal medicine*. 1999; 130(6):515–24. PMID: [10075620](https://pubmed.ncbi.nlm.nih.gov/10075620/).
5. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Annals of internal medicine*. 2006; 144(3):201–9. Epub 2006/02/08. PMID: [16461965](https://pubmed.ncbi.nlm.nih.gov/16461965/).
6. Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. *Journal of clinical epidemiology*. 2008; 61(11):1085–94. Epub 2009/02/12. doi: [10.1016/j.jclinepi.2008.04.008](https://doi.org/10.1016/j.jclinepi.2008.04.008) PMID: [19208371](https://pubmed.ncbi.nlm.nih.gov/19208371/).
7. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *Bmj*. 2009; 338:b605. Epub 2009/05/30. doi: [10.1136/bmj.b605](https://doi.org/10.1136/bmj.b605) PMID: [19477892](https://pubmed.ncbi.nlm.nih.gov/19477892/).
8. Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. *Journal of thrombosis and haemostasis: JTH*. 2013; 11 Suppl 1:129–41. doi: [10.1111/jth.12262](https://doi.org/10.1111/jth.12262) PMID: [23809117](https://pubmed.ncbi.nlm.nih.gov/23809117/).
9. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *Journal of clinical epidemiology*. 2003; 56(9):826–32. PMID: [14505766](https://pubmed.ncbi.nlm.nih.gov/14505766/).
10. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *Journal of clinical epidemiology*. 2003; 56(5):441–7. Epub 2003/06/19. PMID: [12812818](https://pubmed.ncbi.nlm.nih.gov/12812818/).
11. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. New York: Springer; 2009. xxviii, 497 p. p.
12. Wallace E, Smith SM, Perera-Salazar R, Vaucher P, McCowan C, Collins G, et al. Framework for the impact analysis and implementation of Clinical Prediction Rules (CPRs). *BMC medical informatics and decision making*. 2011; 11:62. Epub 2011/10/18. doi: [10.1186/1472-6947-11-62](https://doi.org/10.1186/1472-6947-11-62) PMID: [21999201](https://pubmed.ncbi.nlm.nih.gov/21999201/); PubMed Central PMCID: PMC3216240.
13. Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Annals of emergency medicine*. 1999; 33(4):437–47. PMID: [10092723](https://pubmed.ncbi.nlm.nih.gov/10092723/).
14. Mallett S, Royston P, Dutton S, Waters R, Altman DG. Reporting methods in studies developing prognostic models in cancer: a review. *BMC medicine*. 2010; 8:20. doi: [10.1186/1741-7015-8-20](https://doi.org/10.1186/1741-7015-8-20) PMID: [20353578](https://pubmed.ncbi.nlm.nih.gov/20353578/); PubMed Central PMCID: PMC2856521.
15. Mallett S, Royston P, Waters R, Dutton S, Altman DG. Reporting performance of prognostic models in cancer: a review. *BMC medicine*. 2010; 8:21. doi: [10.1186/1741-7015-8-21](https://doi.org/10.1186/1741-7015-8-21) PMID: [20353579](https://pubmed.ncbi.nlm.nih.gov/20353579/); PubMed Central PMCID: PMC2857810.

16. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC medicine*. 2011; 9:103. doi: [10.1186/1741-7015-9-103](https://doi.org/10.1186/1741-7015-9-103) PMID: [21902820](https://pubmed.ncbi.nlm.nih.gov/21902820/); PubMed Central PMCID: PMC3180398.
17. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS medicine*. 2012; 9(5):1–12. Epub 2012/05/26. doi: [10.1371/journal.pmed.1001221](https://doi.org/10.1371/journal.pmed.1001221) PMID: [22629234](https://pubmed.ncbi.nlm.nih.gov/22629234/); PubMed Central PMCID: PMC3358324.
18. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC medical research methodology*. 2014; 14:40. doi: [10.1186/1471-2288-14-40](https://doi.org/10.1186/1471-2288-14-40) PMID: [24645774](https://pubmed.ncbi.nlm.nih.gov/24645774/); PubMed Central PMCID: PMC3999945.
19. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA: the journal of the American Medical Association*. 1995; 273(5):408–12. PMID: [7823387](https://pubmed.ncbi.nlm.nih.gov/7823387/).
20. Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *Bmj*. 2008; 336(7644):601–5. doi: [10.1136/bmj.39465.451748.AD](https://doi.org/10.1136/bmj.39465.451748.AD) PMID: [18316340](https://pubmed.ncbi.nlm.nih.gov/18316340/); PubMed Central PMCID: PMC2267990.
21. Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schunemann H, et al. Randomisation to protect against selection bias in healthcare trials. *The Cochrane database of systematic reviews*. 2011;(4):MR000012. doi: [10.1002/14651858.MR000012.pub3](https://doi.org/10.1002/14651858.MR000012.pub3) PMID: [21491415](https://pubmed.ncbi.nlm.nih.gov/21491415/).
22. Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Annals of internal medicine*. 2012; 157(6):429–38. doi: [10.7326/0003-4819-157-6-201209180-00537](https://doi.org/10.7326/0003-4819-157-6-201209180-00537) PMID: [22945832](https://pubmed.ncbi.nlm.nih.gov/22945832/).
23. Lijmer JG, Mol BW, Heisterkamp S, Bossel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA: the journal of the American Medical Association*. 1999; 282(11):1061–6. PMID: [10493205](https://pubmed.ncbi.nlm.nih.gov/10493205/).
24. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne*. 2006; 174(4):469–76. doi: [10.1503/cmaj.050090](https://doi.org/10.1503/cmaj.050090) PMID: [16477057](https://pubmed.ncbi.nlm.nih.gov/16477057/); PubMed Central PMCID: PMC1373751.
25. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of internal medicine*. 2003; 138(1):W1–12. PMID: [12513067](https://pubmed.ncbi.nlm.nih.gov/12513067/).
26. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC medical research methodology*. 2003; 3:25. doi: [10.1186/1471-2288-3-25](https://doi.org/10.1186/1471-2288-3-25) PMID: [14606960](https://pubmed.ncbi.nlm.nih.gov/14606960/); PubMed Central PMCID: PMC305345.
27. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine*. 2011; 155(8):529–36. doi: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009) PMID: [22007046](https://pubmed.ncbi.nlm.nih.gov/22007046/).
28. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *Journal of clinical epidemiology*. 2005; 58(5):475–83. Epub 2005/04/23. doi: [10.1016/j.jclinepi.2004.06.017](https://doi.org/10.1016/j.jclinepi.2004.06.017) PMID: [15845334](https://pubmed.ncbi.nlm.nih.gov/15845334/).
29. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med*. 2004; 140(3):189–202. Epub 2004/02/06. PMID: [14757617](https://pubmed.ncbi.nlm.nih.gov/14757617/).
30. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Group Q-S. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol*. 2013; 66(10):1093–104. doi: [10.1016/j.jclinepi.2013.05.014](https://doi.org/10.1016/j.jclinepi.2013.05.014) PMID: [23958378](https://pubmed.ncbi.nlm.nih.gov/23958378/)
31. Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. *Clin Chem*. 2005; 51(8):1335–41. doi: [10.1373/clinchem.2005.048595](https://doi.org/10.1373/clinchem.2005.048595) PMID: [15961549](https://pubmed.ncbi.nlm.nih.gov/15961549/).
32. Wong SS, Wilczynski NL, Haynes RB, Ramkissoon Singh R, Hedges T. Developing optimal search strategies for detecting sound clinical prediction studies in MEDLINE. *AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium*. 2003:728–32. PMID: [14728269](https://pubmed.ncbi.nlm.nih.gov/14728269/); PubMed Central PMCID: PMC1479983.
33. Holland JL, Wilczynski NL, Haynes RB, Hedges T. Optimal search strategies for identifying sound clinical prediction studies in EMBASE. *BMC medical informatics and decision making*. 2005; 5:11. doi: [10.1186/1472-6947-5-11](https://doi.org/10.1186/1472-6947-5-11) PMID: [15862125](https://pubmed.ncbi.nlm.nih.gov/15862125/); PubMed Central PMCID: PMC1097733.

34. Montori VM, Wilczynski NL, Morgan D, Haynes RB, Hedges T. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *Bmj*. 2005; 330(7482):68. doi: [10.1136/bmj.38336.804167.47](https://doi.org/10.1136/bmj.38336.804167.47) PMID: [15619601](https://pubmed.ncbi.nlm.nih.gov/15619601/); PubMed Central PMCID: PMC543864.
35. Wilczynski NL, Haynes RB, Hedges T. EMBASE search strategies achieved high sensitivity and specificity for retrieving methodologically sound systematic reviews. *Journal of clinical epidemiology*. 2007; 60(1):29–33. doi: [10.1016/j.jclinepi.2006.04.001](https://doi.org/10.1016/j.jclinepi.2006.04.001) PMID: [17161751](https://pubmed.ncbi.nlm.nih.gov/17161751/).
36. Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Statistics in medicine*. 2002; 21(11):1513–24. doi: [10.1002/sim.1184](https://doi.org/10.1002/sim.1184) PMID: [12111917](https://pubmed.ncbi.nlm.nih.gov/12111917/).
37. Zamora J, Abaira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC medical research methodology*. 2006; 6:31. doi: [10.1186/1471-2288-6-31](https://doi.org/10.1186/1471-2288-6-31) PMID: [16836745](https://pubmed.ncbi.nlm.nih.gov/16836745/); PubMed Central PMCID: PMC1552081.
38. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Medical decision making: an international journal of the Society for Medical Decision Making*. 1993; 13(4):313–21. PMID: [8246704](https://pubmed.ncbi.nlm.nih.gov/8246704/).
39. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Statistics in medicine*. 1993; 12(14):1293–316. PMID: [8210827](https://pubmed.ncbi.nlm.nih.gov/8210827/).
40. StataCorp. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP; 2011.
41. Kontopantelis E, Reeves D. metaan: Random-effects meta-analysis. *Stata J*. 2010; 10(3):395–407. PMID: [WOS:000282853700007](https://pubmed.ncbi.nlm.nih.gov/WOS:000282853700007/).
42. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *Journal of general internal medicine*. 2007; 22(11):1596–602. Epub 2007/09/18. doi: [10.1007/s11606-007-0333-y](https://doi.org/10.1007/s11606-007-0333-y) PMID: [17874169](https://pubmed.ncbi.nlm.nih.gov/17874169/); PubMed Central PMCID: PMC2219806.
43. Warnick EM, Bracken MB, Kasl S. Screening efficiency of the child behavior checklist and strengths and difficulties questionnaire: A systematic review. *Child and Adolescent Mental Health*. 2008; 13(3):140–7. doi: <http://dx.doi.org/10.1111/j.1475-3588.2007.00461.x> PMID: [2008386644](https://pubmed.ncbi.nlm.nih.gov/2008386644/).
44. Banal F, Dougados M, Combescur C, Gossec L. Sensitivity and specificity of the American College of Rheumatology 1987 criteria for the diagnosis of rheumatoid arthritis according to disease duration: a systematic literature review and meta-analysis. *Annals of the rheumatic diseases*. 2009; 68(7):1184–91. Epub 2008/08/30. doi: [10.1136/ard.2008.093187](https://doi.org/10.1136/ard.2008.093187) PMID: [18728049](https://pubmed.ncbi.nlm.nih.gov/18728049/).
45. Dowling S, Spooner CH, Liang Y, Dryden DM, Friesen C, Klassen TP, et al. Accuracy of Ottawa Ankle Rules to exclude fractures of the ankle and midfoot in children: a meta-analysis. *Academic emergency medicine: official journal of the Society for Academic Emergency Medicine*. 2009; 16(4):277–87. Epub 2009/02/04. doi: [10.1111/j.1553-2712.2008.00333.x](https://doi.org/10.1111/j.1553-2712.2008.00333.x) PMID: [19187397](https://pubmed.ncbi.nlm.nih.gov/19187397/).
46. Geomini P, Kruitwagen R, Bremer GL, Cnossen J, Mol BW. The accuracy of risk scores in predicting ovarian malignancy: a systematic review. *Obstetrics and gynecology*. 2009; 113(2 Pt 1):384–94. Epub 2009/01/22. doi: [10.1097/AOG.0b013e318195ad17](https://doi.org/10.1097/AOG.0b013e318195ad17) PMID: [19155910](https://pubmed.ncbi.nlm.nih.gov/19155910/).
47. Hewitt C, Gilbody S, Brealey S, Paulden M, Palmer S, Mann R, et al. Methods to identify postnatal depression in primary care: an integrated evidence synthesis and value of information analysis. *Health Technol Assess*. 2009; 13(36):1–145, 7-230. Epub 2009/07/25. doi: [10.3310/hta13360](https://doi.org/10.3310/hta13360) PMID: [19624978](https://pubmed.ncbi.nlm.nih.gov/19624978/).
48. Brennan C, Worrall-Davies A, McMillan D, Gilbody S, House A. The Hospital Anxiety and Depression Scale: a diagnostic meta-analysis of case-finding ability. *Journal of psychosomatic research*. 2010; 69(4):371–8. Epub 2010/09/18. doi: [10.1016/j.jpsychores.2010.04.006](https://doi.org/10.1016/j.jpsychores.2010.04.006) PMID: [20846538](https://pubmed.ncbi.nlm.nih.gov/20846538/).
49. Ceriani E, Combescur C, Le Gal G, Nendaz M, Perneger T, Bounameaux H, et al. Clinical prediction rules for pulmonary embolism: a systematic review and meta-analysis. *Journal of thrombosis and haemostasis: JTH*. 2010; 8(5):957–70. Epub 2010/02/13. doi: [10.1111/j.1538-7836.2010.03801.x](https://doi.org/10.1111/j.1538-7836.2010.03801.x) PMID: [20149072](https://pubmed.ncbi.nlm.nih.gov/20149072/).
50. Giles MF, Rothwell PM. Systematic review and pooled analysis of published and unpublished validations of the ABCD and ABCD2 transient ischemic attack risk scores. *Stroke; a journal of cerebral circulation*. 2010; 41(4):667–73. Epub 2010/02/27. doi: [10.1161/STROKEAHA.109.571174](https://doi.org/10.1161/STROKEAHA.109.571174) PMID: [20185786](https://pubmed.ncbi.nlm.nih.gov/20185786/).
51. Harrington L, Luquire R, Vish N, Winter M, Wilder C, Houser B, et al. Meta-analysis of fall-risk tools in hospitalized adults. *The Journal of nursing administration*. 2010; 40(11):483–8. Epub 2010/10/28. doi: [10.1097/NNA.0b013e3181f88fbd](https://doi.org/10.1097/NNA.0b013e3181f88fbd) PMID: [20978417](https://pubmed.ncbi.nlm.nih.gov/20978417/).
52. McPhail MJ, Wendon JA, Bernal W. Meta-analysis of performance of Kings's College Hospital Criteria in prediction of outcome in non-paracetamol-induced acute liver failure. *Journal of hepatology*. 2010; 53(3):492–9. Epub 2010/06/29. doi: [10.1016/j.jhep.2010.03.023](https://doi.org/10.1016/j.jhep.2010.03.023) PMID: [20580460](https://pubmed.ncbi.nlm.nih.gov/20580460/).

53. Mitchell AJ, Bird V, Rizzo M, Meader N. Diagnostic validity and added value of the Geriatric Depression Scale for depression in primary care: a meta-analysis of GDS30 and GDS15. *Journal of affective disorders*. 2010; 125(1-3):10–7. Epub 2009/10/06. doi: [10.1016/j.jad.2009.08.019](https://doi.org/10.1016/j.jad.2009.08.019) PMID: [19800132](https://pubmed.ncbi.nlm.nih.gov/19800132/).
54. Mitchell AJ, Bird V, Rizzo M, Meader N. Which version of the geriatric depression scale is most useful in medical settings and nursing homes? Diagnostic validity meta-analysis. *The American journal of geriatric psychiatry: official journal of the American Association for Geriatric Psychiatry*. 2010; 18(12):1066–77. Epub 2010/12/16. PMID: [21155144](https://pubmed.ncbi.nlm.nih.gov/21155144/).
55. Mitchell AJ, Meader N, Symonds P. Diagnostic validity of the Hospital Anxiety and Depression Scale (HADS) in cancer and palliative settings: a meta-analysis. *Journal of affective disorders*. 2010; 126(3):335–48. doi: [10.1016/j.jad.2010.01.067](https://doi.org/10.1016/j.jad.2010.01.067) PMID: [20207007](https://pubmed.ncbi.nlm.nih.gov/20207007/).
56. Serrano LA, Hess EP, Bellolio MF, Murad MH, Montori VM, Erwin PJ, et al. Accuracy and quality of clinical decision rules for syncope in the emergency department: a systematic review and meta-analysis. *Annals of emergency medicine*. 2010; 56(4):362–73 e1. Epub 2010/09/28. doi: [10.1016/j.annemergmed.2010.05.013](https://doi.org/10.1016/j.annemergmed.2010.05.013) PMID: [20868906](https://pubmed.ncbi.nlm.nih.gov/20868906/); PubMed Central PMCID: PMC2946941.
57. Stengel D, Bauwens K, Rademacher G, Mutze S, Ekkernkamp A. Association between compliance with methodological standards of diagnostic research and reported test accuracy: meta-analysis of focused assessment of US for trauma. *Radiology*. 2005; 236(1):102–11. doi: [10.1148/radiol.2361040791](https://doi.org/10.1148/radiol.2361040791) PMID: [15983072](https://pubmed.ncbi.nlm.nih.gov/15983072/).
58. Hulsemann JL, Zeidler H. Diagnostic evaluation of classification criteria for rheumatoid arthritis and reactive arthritis in an early synovitis outpatient clinic. *Annals of the rheumatic diseases*. 1999; 58(5):278–80. PMID: [10225811](https://pubmed.ncbi.nlm.nih.gov/10225811/); PubMed Central PMCID: PMC1752884.
59. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Annals of internal medicine*. 2015; 162(1):55–63. doi: [10.7326/M14-0697](https://doi.org/10.7326/M14-0697) PMID: [25560714](https://pubmed.ncbi.nlm.nih.gov/25560714/).