# MAXIMIZING RESEARCH VALUE:

## ADEQUATE REPORTING AND EFFECTIVE (DE-)IMPLEMENTATION STRATEGIES

**Pauline Heus**

# Maximizing research value: adequate reporting and effective (de-)implementation strategies

Pauline Heus

Maximizing research value: adequate reporting and effective (de-)implementation strategies
Pauline Heus
PhD thesis, with a summary in Dutch
Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, the Netherlands

ISBN: 978-94-6375-925-0

Cover image “Het hele spectrum in zicht” by Madeleine Metsaars
Lay-out by Lara Leijtens, persoonlijkproefschrift.nl
Printed by Ridderprint | www.ridderprint.nl

# Maximizing research value: adequate reporting and effective (de-)implementation strategies

**Maximaliseren van de waarde van onderzoek: adequate rapportage en effectieve (de-)implementatiestrategieën**

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

dinsdag 30 juni 2020 des middags te 2.30 uur

door

Pauline Heus

geboren op 28 november 1978
te Heemskerk

Promotoren:

Prof. dr. K.G.M. Moons

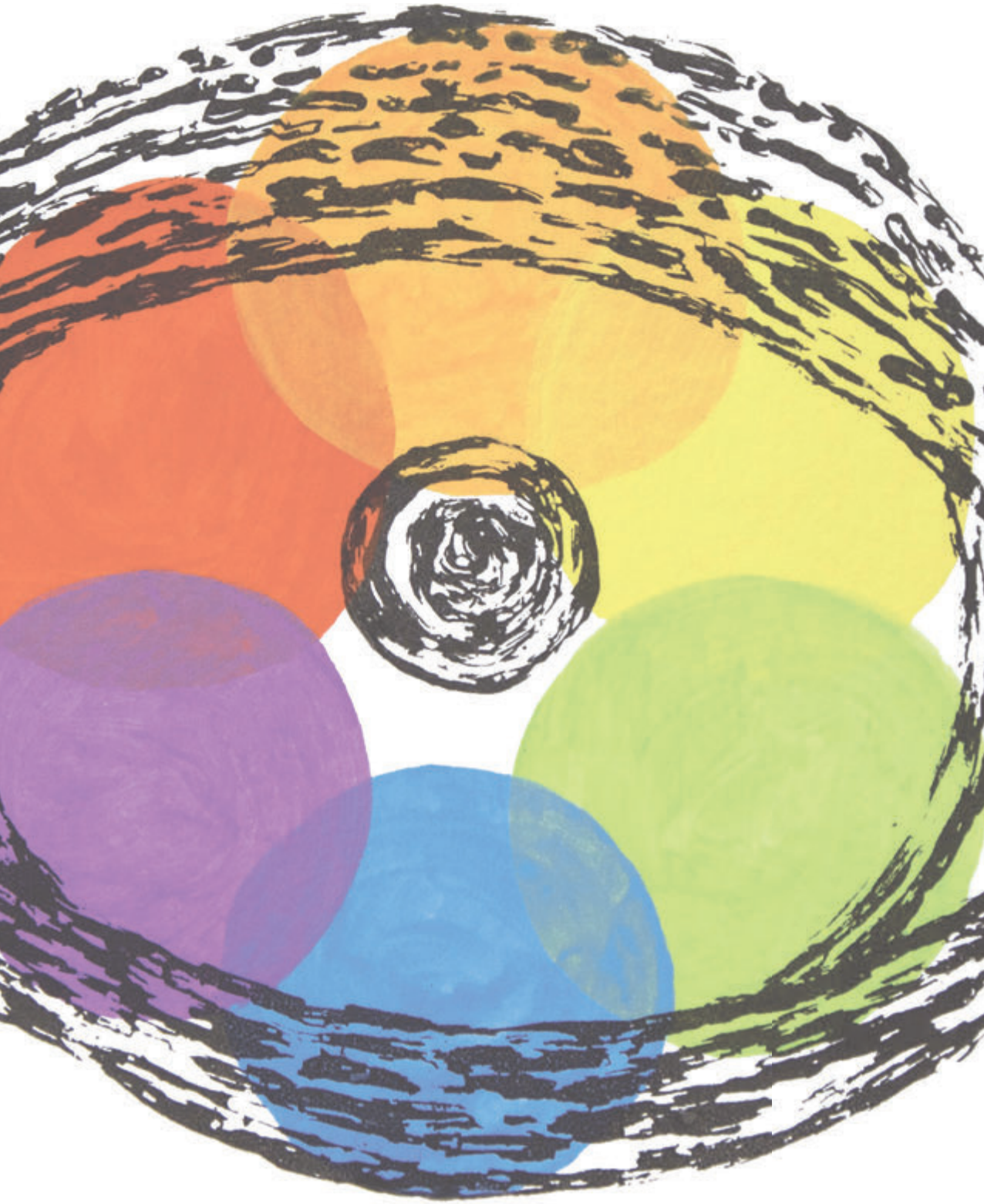
Prof. dr. R.J.P.M. Scholten

Copromotoren:

Dr. L. Hooft

Dr. J.B. Reitsma

## Contents



# Chapter 1

## General introduction

On several occasions during the past few years I experienced the healthcare system as a patient or patient's relative. Such situations usually give rise to a number of questions and decisions to make regarding diagnosis, treatment, and prognosis. Everyone who has been in that position recognizes the value of clear and consistent information, whether coming from healthcare providers, found in clinical guidelines or on the internet. Inconsistent information and practice variation in times of (possible) disease is undesirable. The principles of evidence-based medicine contribute to reducing this unwanted variation and insecurity.

The term evidence-based medicine (EBM) was first introduced in the early nineties of the past century.[1,2] It is defined as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients".[3] EBM describes the process of integrating the available information from clinical research (evidence) with clinical expertise and patient preferences. Adding evidence to the decision making process reduces the influence of just clinical expertise and intuition, which had been the main drivers for clinical decision making before the 1990's.

The essential first step in the EBM process is the generation of evidence, in which the appropriate research designs, methods, and analyses, should be used to answer questions relevant to users of research. Next, availability and accessibility of evidence are prerequisites for applying evidence in clinical practice. Evidence dissemination starts by researchers writing useful reports of their research. Research reports should be a complete, accurate and transparent reflection of the research performed to enable critical appraisal of the applied methods, unambiguous interpretation of the research results and applicability of these results in clinical care. Without a clear description of the research question addressed, the methods used, the resulting data and the potential implications of these findings, the usability of research is reduced and the research efforts may be considered as wasted.[4]

Reporting guidelines have been developed to guide authors and improve the reporting of research. These guidelines are developed according to a specific methodology and can come in the form of checklists, flow diagrams or structured texts. The Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network is an international collaboration that promotes responsible reporting of health research by providing resources and training, and by assisting in reporting guideline development, dissemination, and implementation.[5,6] As reporting guidelines are usually developed for a specific type of research, many reporting guidelines exist for the various types of study designs. Well-known examples are the CONsolidated

Standards Of Reporting Trials (CONSORT) statement, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, Strengthening the Reporting of OBservational studies in Epidemiology (STROBE) statement, and STAndards for Reporting of Diagnostic Accuracy (STARD) statement, and the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis(TRIPOD) statement, which address the reporting of randomized trials, systematic reviews and meta-analyses, observational studies, diagnostic accuracy studies, and prediction model studies, respectively.[7-11]



Although important, transparent and accurate reporting does not guarantee that research evidence will be used in practice. Additional active and targeted strategies are often needed to achieve the uptake of evidence into routine practice (implementation). Implementation requires behavioural change, which is influenced by factors related to the individual healthcare provider or patient (e.g. beliefs, experiences, motivation), as well as contextual factors at the social, organisational or wider environmental level (e.g. time, resources,) factors.[12-14] Implementation science in the healthcare sector focusses on the methods to enhance the uptake of evidence in clinical practices to improve the quality and effectiveness of health services.[15] It has two main fields of interest: (1) identification of barriers and facilitators to uptake of medical research evidence across various contextual levels (like patients, providers, organization, and other stakeholders); (2) development and application of strategies to overcome these barriers and enhance the facilitators.[16,17] Many theories, models, and frameworks exist that can assist with identifying and classifying barriers and facilitators and the potential interventions to address them.[17-21]

## Aim and outline of this thesis

The aim of this thesis is to explore and improve the methods to report healthcare research and implement research findings (evidence), which thus are both essential components to facilitate EBM. The first part of this thesis focuses on the reporting of prediction model studies and the TRIPOD reporting guideline that aims to improve the adequacy of reporting of this study type (Chapters 2-5). The second part addresses the implementation of evidence that recommends to no longer provide a specific healthcare practice (Chapters 6-8).

### Reporting of prediction model studies

Prediction models can assist in clinical decision making by estimating an individual's probability that a specific outcome or condition is present (diagnostic models) or that a specific outcome or event will occur in the future (prognostic models), based on

multiple pieces of information of that individual.[22] Studies about prediction models may address the development of a new model, validation of an existing, previously developed model in other individuals, and the evaluation of an existing model's extension or updating.[23-26] For this type of study the TRIPOD statement was published in 2015.[11,27]

We evaluated the completeness of reporting of prediction model studies, published just before the introduction of the TRIPOD statement, which is described in **Chapter 2**. For this assessment we transformed the original 22 items of the TRIPOD statement into a systematic and transparent adherence assessment form. In **Chapter 3** we share our experiences with designing this form and creating TRIPOD adherence scoring rules. We present the development of additional guidance for reporting prediction model studies in journal and conference abstracts in **Chapter 4**. **Chapter 5** describes the endorsement of TRIPOD and other reporting guidelines by medical journals. In addition, this chapter reports on an online survey among journal editors to identify potential barriers and facilitators to the implementation of reporting guidelines.

**Implementing evidence to no longer provide a specific healthcare practice**

Implementation science and implementing evidence is not always directed to using a new intervention or healthcare practice, evidence can also include recommendations to stop specific interventions or healthcare practices that are currently used in daily practice. Low-value care is the term to describe healthcare practices leading to no or little clinical benefit for the patient, considering the costs, risks, and available alternatives.[28,29] It is closely related to the concept of overuse, including both overtesting and overtreatment. These low-value healthcare practices should be stopped or not routinely be provided. The active process of reducing low-value care is called de-implementation.[30] Like implementation, de-implementation involves changing behavior, however, stopping or changing an existing practice is likely to be more difficult than starting a new one.[31] Interventions to reduce low-value care should address the specific individual and contextual factors relating to the low-value healthcare practices of interest.

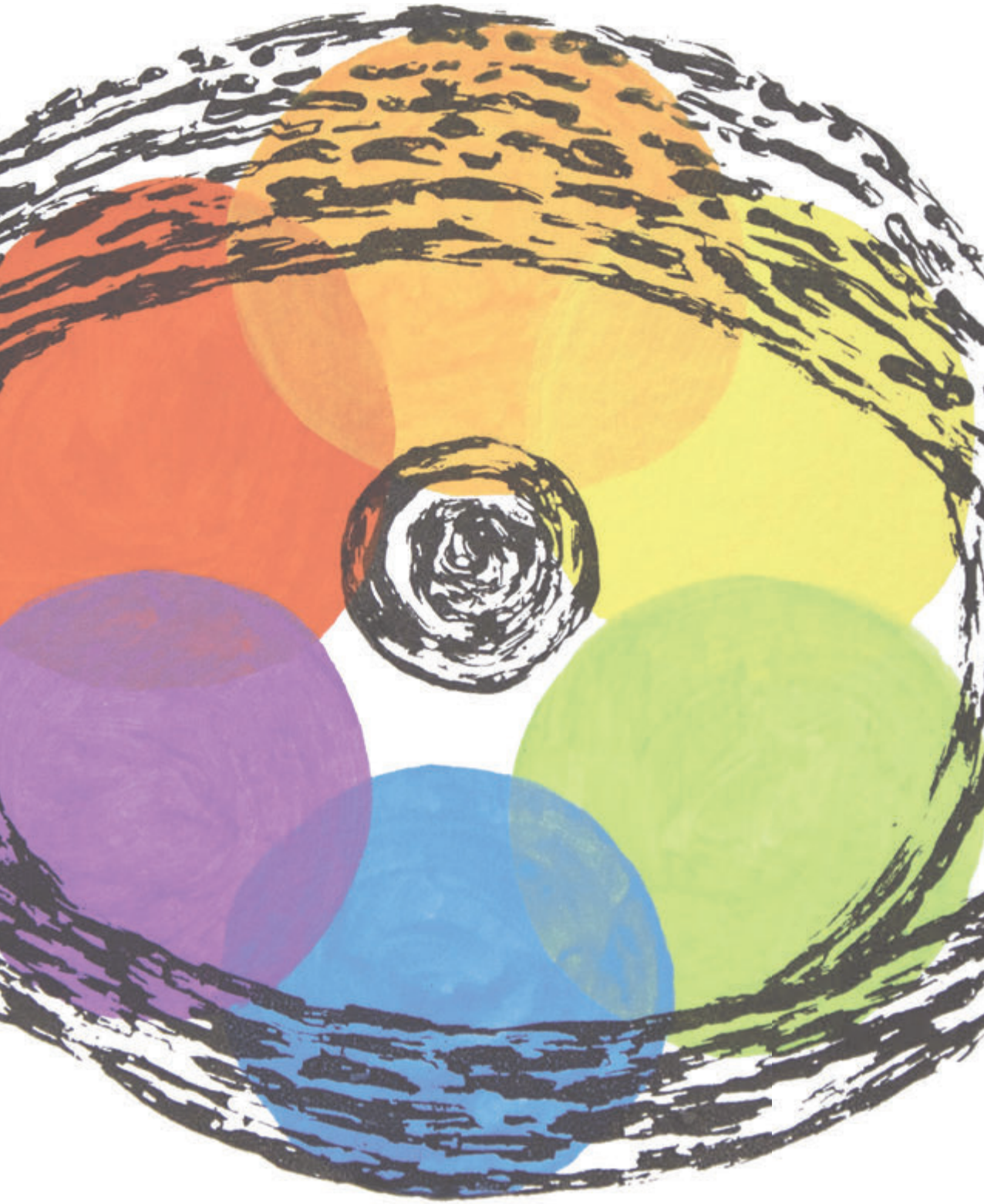
We synthesized the existing evidence regarding potential barriers and facilitators to de-implementation in healthcare settings in **Chapter 6. Chapter 7** describes a systematic review of de-implementation studies in which we compared the effectiveness of various strategies to reduce low-value care and aimed to identify characteristics associated with their success. In **Chapter 8** we focus on reducing the use of low-value medical tests in primary care settings.

This thesis ends in **Chapter 9** with a reflection on the lessons learned and the implications for practice and research. The challenges around de-implementation are illustrated with a case study.

## References


1. Guyatt GH. Evidence-based medicine. *ACP J Club.* 1991;114(2):A16.
2. Guyatt G, Cairns J, Churchill D, Cook D, Haynes B, et al. Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *JAMA.* 1992;268(17):2420-2425.
3. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ.* 1996;312(7023):71-72.
4. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet.* 2014;383(9913):267-276.
5. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. *Open Med.* 2008;2(2):e49-e50.
6. The EQUATOR Network | Enhancing the QUAlity and Transparency Of health Research. https://www.equator-network.org/. Accessed 2 February, 2020.
7. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med.* 2010;152(11):726-732.
8. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097.
9. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med.* 2007;147(8):573-577.
10. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527.
11. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55-63.
12. Haines A, Donald A. Making better use of research findings. *BMJ.* 1998;317(7150):72-75.
13. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA.* 1999;Jama282(15):1458-1465.
14. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet.* 2003;362(9391):1225-1230.
15. Eccles MP, Mittman BS. Welcome to Implementation Science. *Implement Sci.* 2006;1(1):1.
16. Bauer MS, Kirchner J. Implementation science: What is it and why should I care? *Psychiatry Res.* 2020;283:112376.

17. Shelton RC, Lee M, Brotzman LE, Wolfenden L, Nathan N, Wainberg ML. What Is Dissemination and Implementation Science?: An Introduction and Opportunities to Advance Behavioral Medicine and Public Health Globally. *Int J Behav Med*. 2020.
18. Michie S, van Stralen MM, West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implement Sci*. 2011;6(1):42.
19. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. *Qual Saf Health Care*. 2005;14(1):26-33.
20. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci*. 2015;10:21.
21. Wensing M, Grol R. Knowledge translation in health: how implementation science could contribute more. *BMC Med*. 2019;17(1):88.
22. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381.
23. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98(9):683-690.
24. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98(9):691-698.
25. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338:b605.
26. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009;338:b604.
27. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-73.
28. Scott IA, Duckett SJ. In search of professional consensus in defining and reducing low-value care. *Med J Aust*. 2015;203(4):179-181.
29. Colla CH, Mainor AJ, Hargreaves C, Sequist T, Morden N. Interventions Aimed at Reducing Use of Low-Value Health Services: A Systematic Review. *Med Care Res Rev*. 2017;74(5).
30. Niven DJ, Mrklas KJ, Holodinsky JK, Straus SE, Hemmelgarn BR, et al. Towards understanding the de-adoption of low-value clinical practices: A scoping review. *BMC Medicine*. 2015;13.
31. van Bodegom-Vos L, Davidoff F, Marang-van de Mheen PJ. Implementation and de-implementation: two sides of the same coin? *BMJ Qual Saf*. 2016.



# Chapter 2

## Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement

Pauline Heus
Johanna AAG Damen
Romin Pajouheshnia
Rob JPM Scholten
Johannes B Reitsma
Gary S Collins
Douglas G Altman
Karel GM Moons
Lotty Hooft

*BMC Medicine (2018) 16:120*

## Abstract

### Background

As complete reporting is essential to judge the validity and applicability of multivariable prediction models, a guideline for the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) was introduced. We assessed the completeness of reporting of prediction model studies published just before the introduction of the TRIPOD statement, to refine and tailor its implementation strategy.

### Methods

Within each of 37 clinical domains, 10 journals with the highest journal impact factor were selected. A PubMed search was performed to identify prediction model studies published before the launch of TRIPOD (May 2014) in these journals. Eligible publications reported on the development or external validation of a multivariable prediction model (either diagnostic or prognostic), or on the incremental value of adding a predictor to an existing model.

### Results

We included 146 publications (84% prognostic), from which we assessed 170 models: 73 (43%) model development, 43 (25%) external validation, 33 (19%) incremental value, and 21 (12%) combined development and external validation of the same model. Overall, publications adhered to a median of 44% ($25^{th}$–$75^{th}$ percentile: 35% to 52%) of TRIPOD items, with 44% (35% to 53%) for prognostic and 41% (34% to 48%) for diagnostic models. TRIPOD items that were completely reported for less than 25% of the models concerned abstract (2%), title (5%), blinding of predictor assessment (6%), comparison of development and validation data (11%), model updating (14%), model performance (14%), model specification (17%), characteristics of participants (21%), model performance measures (methods) (21%), and model building procedures (24%). Most often reported were TRIPOD items regarding overall interpretation (96%), source of data (95%), and risk groups (90%).

### Conclusions

More than half of the items considered essential for transparent reporting were not fully addressed in publications of multivariable prediction model studies. Essential information for using a model in individual risk prediction, i.e. model specifications and model performance, was incomplete for over 80% of the models. Items that require improved reporting are title, abstract, and model building procedures, as they are crucial for identification and external validation of prediction models.

## Background

Multivariable prediction models (risk scores or prediction rules) estimate an individual's probability or risk that a specific disease or condition is present (diagnostic models) or that a specific event will occur in the future (prognostic models) based on multiple characteristics or pieces of information of that individual.[1] Such models are increasingly used by healthcare providers to support clinical decision making or to inform patients or relatives. Studies about prediction models may address the development of a new model, validation of an existing, previously developed model in other individuals (with or without adjusting or updating the model to the validation setting), or a combination of these two.[2-5] Some prediction model studies evaluate the addition of a single predictor to an existing model (incremental value).[4]



In addition to appropriate design, conduct and analysis, reporting of prediction model studies should be complete and accurate. Complete reporting of research facilitates study replication, assessment of the study validity (risk of bias), interpretation of the results, and judgement of applicability of the study results (e.g. the prediction model itself) to other individuals or settings. Clinicians and other stakeholders can only use previously developed and validated prediction models when all relevant information is available for calculating predicted risks at an individual level. High quality information about prediction model studies is therefore essential.

Previous systematic reviews showed that within different clinical domains the quality of reporting of prediction models is suboptimal.[6-11] To improve the reporting of studies of prediction models, a guideline for the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) was launched in January 2015 in over 10 medical journals.[12,13] The TRIPOD statement is a checklist of 22 items considered essential for informative reporting of prediction model studies. Both diagnostic and prognostic prediction model studies are covered by the TRIPOD statement, and the checklist can be used for all types of prediction model studies (development, external validation, and incremental value) within all clinical domains.

In this comprehensive literature review, we assessed the completeness of reporting of prediction model studies that were published just before the introduction of the TRIPOD statement. Our results provide key clues to further refine and tailor the implementation strategy of the TRIPOD statement.

## Methods

### Identification of prediction model studies

To cover a wide range of clinical domains we started with 37 subject categories (2012 Journal Citation Reports®)[14] from which we selected the 10 journals with the highest Journal Impact Factor (Additional file 1). After deduplication, 341 unique journals remained. We performed a search in PubMed to identify prediction model studies published in these journals before the launch of TRIPOD (May 2014), using a validated search filter for identifying prognostic and diagnostic prediction studies (Additional file 2).[15]

Eligible publications described the development or external validation of a multivariable prediction model (either diagnostic or prognostic), or evaluated the incremental value of adding a predictor to an existing model.[1-5,16] We excluded so-called prognostic factor or predictor finding studies, as well as studies evaluating the impact of the use of a prediction model on management or patient outcomes.[3,7,17] We excluded prediction model studies using non-regression techniques (e.g. classification trees, neural networks and machine learning) or pharmacokinetic models. Titles and abstracts of the retrieved publications were screened by one of two authors (JAAGD or PH). After reading the full text report, they judged whether to include or exclude a potentially eligible publication. Any doubts regarding definitive eligibility were discussed, if necessary, with a third author. If we were not able to retrieve the full text of a publication via our institutions, it was excluded.

### Data-extraction

For each included publication we recorded the journal impact factor (2012 Journal Citation Reports®)[14], clinical domain, and whether the purpose of prediction was diagnostic or prognostic. Furthermore, we classified publications into four types of prediction model studies: development, external validation, incremental value, or combination of development and external validation of the same model. A publication could be categorized as more than one type of prediction model study. For example, if a publication reported on both development and external validation, but of different models, it was classified as development as well as external validation. If a publication included multiple prediction model studies of the same type, e.g. two models were developed, we extracted data for only one model. If there was no primary model, we used the model that was studied in the largest sample. Information about study design, sample size, number of predictors in the final model, and predicted outcome was extracted for all included prediction models.

To judge the completeness of the reporting, we transformed items of the TRIPOD statement (Box 1) into a data-extraction form, which was piloted extensively to ensure consistent extraction of the data. The TRIPOD statement consists of 22 main items, of which ten are divided in two (items 3, 4, 6, 7, 14, 15, and 19), three (items 5 and 13), or five (item 10) sub items.[12,13] For TRIPOD items (main or sub items, hereafter just called items) containing multiple reporting elements we extracted information regarding each of these elements. For example, for item 4b *"Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up."* we used three data extraction elements to record information regarding 1) the start of accrual, 2) end of accrual, and 3) end of follow-up. The data extraction form including all data extraction elements can be found on the website of the TRIPOD statement (www.tripod-statement.org ).



**Box 1. Items of the TRIPOD statement**

***Title and abstract***

1. **Title (D; V)**: identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.
2. **Abstract (D; V)**: provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.

***Introduction***

3. **Background and objectives**:
   - **a. (D; V)** Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.
   - **b. (D; V)** Specify the objectives, including whether the study describes the development or validation of the model or both.

***Methods***

4. **Source of data**:
   - **a. (D; V)** Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.
   - **b. (D; V)** Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.
5. **Participants**:
   - **a. (D; V)** Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.
   - **b. (D; V)** Describe eligibility criteria for participants.
   - **c. (D; V)** Give details of treatments received, if relevant.
6. **Outcome**:
   - **a. (D; V)** Clearly define the outcome that is predicted by the prediction model, including how and when assessed.
   - **b. (D; V)** Report any actions to blind assessment of the outcome to be predicted.

7. **Predictors**:
   a. **(D; V)** Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.
   b. **(D; V)** Report any actions to blind assessment of predictors for the outcome and other predictors.
8. **Sample size (D; V)**: explain how the study size was arrived at.
9. **Missing data (D; V)**: Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.
10. **Statistical analysis methods**:
    a. **(D)** Describe how predictors were handled in the analyses.
    b. **(D)** Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.
    c. **(V)**For validation, describe how the predictions were calculated.
    d. **(D; V)** Specify all measures used to assess model performance and, if relevant, to compare multiple models.
    e. **(V)** Describe any model updating (e.g., recalibration) arising from the validation, if done.
11. **Risk groups (D; V)**: Provide details on how risk groups were created, if done.
12. **Development vs. validation (V)**: for validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.

***Results***

13. **Participants**:
    a. **(D; V)** Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.
    b. **(D; V)** Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.
    c. **(V)** For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).
14. **Model development**:
    a. **(D)** Specify the number of participants and outcome events in each analysis.
    b. **(D)**If done, report the unadjusted association between each candidate predictor and outcome.
15. **Model specification**:
    a. **(D)** Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).
    b. **(D)** Explain how to the use the prediction model.
16. **Model performance (D;V)**: report performance measures (with CIs) for the prediction model.
17. **Model-updating (V)**: if done, report the results from any model updating (i.e., model specification, model performance).

*Discussion*

18. **Limitations (D;V)**: discuss any limitations of the study (such as non-representative sample, few events per predictor, missing data).
19. **Interpretation**:
    a. **(V)** For validation, discuss the results with reference to performance in the development data, and any other validation data.
    b. **(D;V)**Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.
20. **Implications (D;V)**: discuss the potential clinical use of the model and implications for future research.

*Other information*

21. **Supplementary information (D;V)**: provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.
22. **Funding (D;V)**: give the source of funding and the role of the funders for the present study.

*D;V*: item relevant to both development and external validation; *D*: item only relevant to development; *V*: item only relevant to external validation

For each data extraction element we judged whether the requested information was available in the publication. If a publication reported both the development and external validation of the same prediction model, we extracted data on the reporting of either separately, and subsequently combined the extracted information for each data extraction element.

Three authors extracted data (JAAGD, PH, RP). If the authors disagreed or were unsure about the reporting of a data extraction element, it was discussed in consensus meetings with the other co-authors.

**Analyses**

Based on the extracted data elements, we first determined whether the reporting of each TRIPOD item was complete (definition see below). We then calculated overall scores for completeness of reporting per model, per publication, and per item of the TRIPOD statement (across models).

*Completeness of reporting of each TRIPOD item*

The reporting of a TRIPOD item was judged to be complete if the requested information for all elements of that particular TRIPOD item was present. For elements belonging to TRIPOD items 4b, 5a, 6a, and 7a we considered a reference to information in another article acceptable. If an element was not applicable to a specific model, for example follow-up might be not relevant in a diagnostic prediction model study

(item 4b), or blinding was a non-issue (e.g. if the predicted outcome was for example overall mortality) (items 6b and 7b), this element was regarded as being reported.

*Overall completeness of reporting per model*
To calculate overall completeness of reporting for each included model we divided the number of completely reported TRIPOD items by the total number of TRIPOD items for that model. The total number of TRIPOD items varies per type of prediction model study, as six of the TRIPOD items only apply to development of a prediction model (10a, 10b, 14a, 14b, 15a, and 15b) and six only to external validation (10c, 10e, 12, 13c, 17, and 19a). This resulted in a total number of 31 TRIPOD items for the reporting of either development or external validation of a prediction model, 37 for the combined reporting of development and external validation of the same prediction model, and 36 for reporting incremental value.

Five items of the TRIPOD statement include an 'if done' or 'if applicable' statement (items 5c, 10e, 11, 14b and 17). If we considered such an item not applicable for a particular study, it was excluded when calculating the completeness of reporting (both in numerator and denominator). Furthermore, item 21 of the TRIPOD statement was excluded from all calculations, as it refers to whether supplementary material was provided.

*Overall completeness of reporting per publication*
The overall reporting per publication equals the reporting per model (see previous paragraph) for publications classified as either development, external validation, incremental value, or combined development and external validation of the same model. For publications classified as more than one type of prediction model study, for example development of a model and external validation of a different model, we combined the reporting of the different prediction model types within that publication. Reporting was considered complete when the reporting of the different types of prediction model studies was complete, except for TRIPOD items 3a and 18-20, for which complete reporting for either type was considered sufficient.

We used linear regression to investigate possible relationships between completeness of reporting per publication as dependent variable, and sample size, journal impact factor, number of predictors in the final model, and prospective study design (as dichotomous variable, yes/no) as independent variables.

*Overall completeness of reporting per item of the TRIPOD statement*

We assessed the overall completeness of reporting of individual items of the TRIPOD statement by dividing the number of models with complete reporting of a particular TRIPOD item by the total number of models in which that item was applicable.

## Results

We included a total of 146 publications (Figure 1). Most publications (122 [84%]) reported prognostic models. From the 146 publications we scored the reporting of 170 prediction models: 73 (43%) concerned model development, 43 (25%) external validation of an existing model, 33 (19%) incremental value of adding a predictor to a model, and 21 (12%) a combination of development and external validation of the same model.



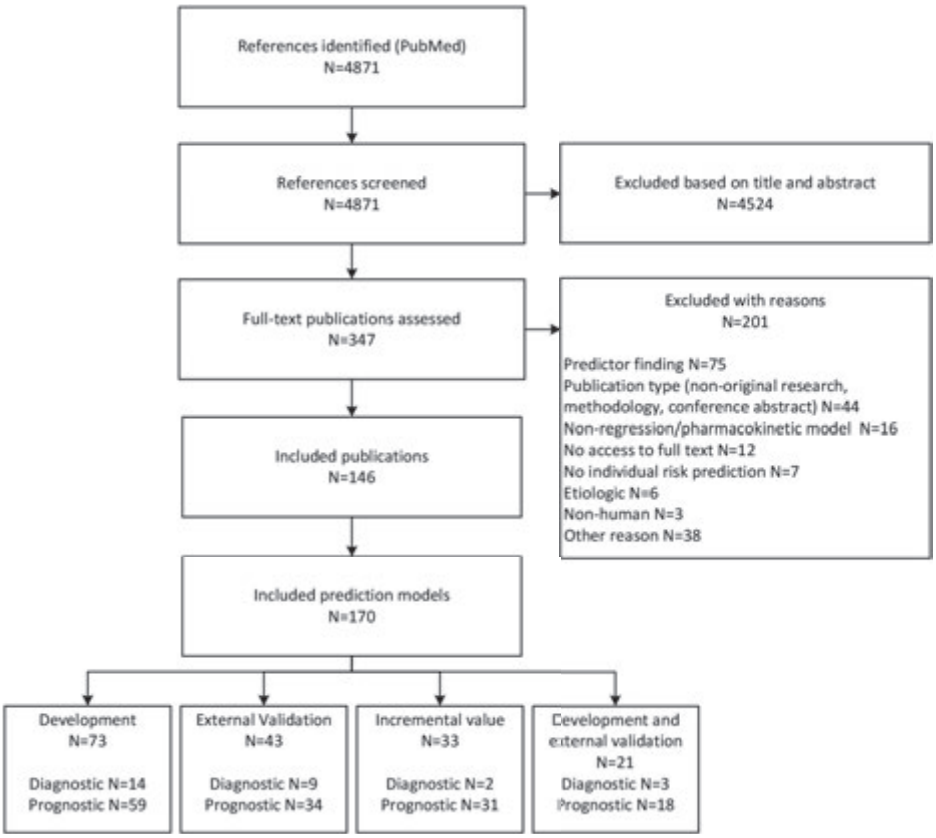



**Figure 1. Flow diagram of selection procedure**

The three clinical domains with most publications of prediction models were critical care medicine (18 [11%]), obstetrics and gynaecology (15 [9%]), and gastroenterology and hepatology (12 [7%]). The median journal impact factor of the publications was 5.3 ($25^{th}$-$75^{th}$ percentile [$P_{25}$-$P_{75}$]: 4.0-7.1). Median sample size of the populations in which a model was studied was 450 ($P_{25}$-$P_{75}$: 200-2005). In the final models a median of 5 ($P_{25}$-$P_{75}$: 3-8) predictors were included and in 23 models (16%) all-cause mortality was the predicted outcome.

**Completeness of reporting per publication**

Overall, publications adhered to between 16% to 81% of the items of the TRIPOD statement with a median of 44% ($P_{25}$-$P_{75}$: 35%-52%) (Figure 2). The reporting quality for prognostic and diagnostic prediction models was comparable, with median adherence of 44% ($P_{25}$-$P_{75}$: 35%-53%) and 41% ($P_{25}$-$P_{75}$: 34%-48%), respectively. The most complete reporting was seen for the combined reporting of development and external validation of the same model (47%, $P_{25}$-$P_{75}$: 35%-54%), followed by the reporting of model development (43%; $P_{25}$-$P_{75}$: 35%-53%), external validation (43%; $P_{25}$-$P_{75}$: 37%-54%), and incremental value (38%; $P_{25}$-$P_{75}$: 33%-49%). No associations were found between completeness of reporting and sample size, journal impact factor, number of predictors in the final model, and prospective study design (data not shown).

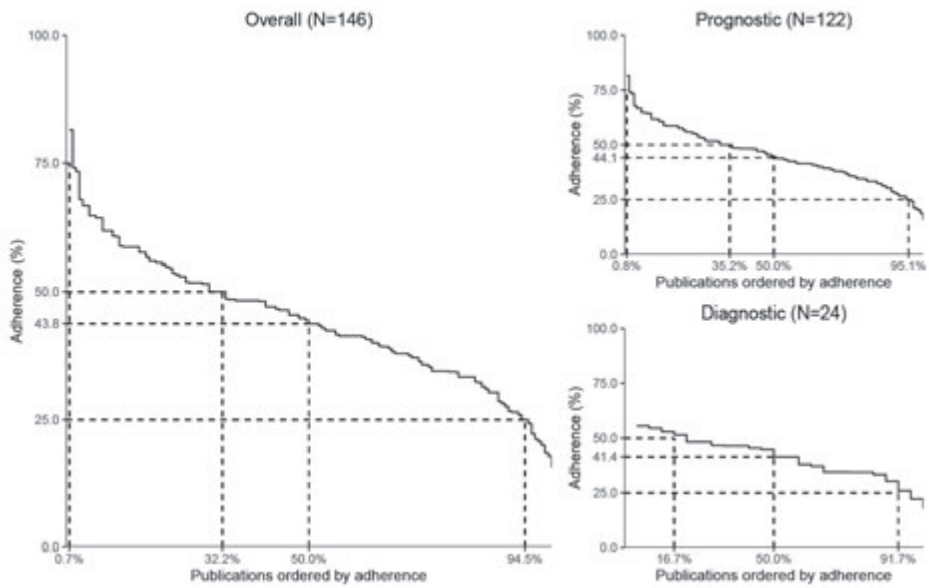



**Figure 2. Reporting across publications: adherence to items of the TRIPOD statement**

### Reporting of individual TRIPOD items

Six TRIPOD items were reported in 75% or more of the 170 models, and 10 items in less than 25% (Table 1).

**Table 1. Completeness of reporting of individual TRIPOD items (n=170 models)**

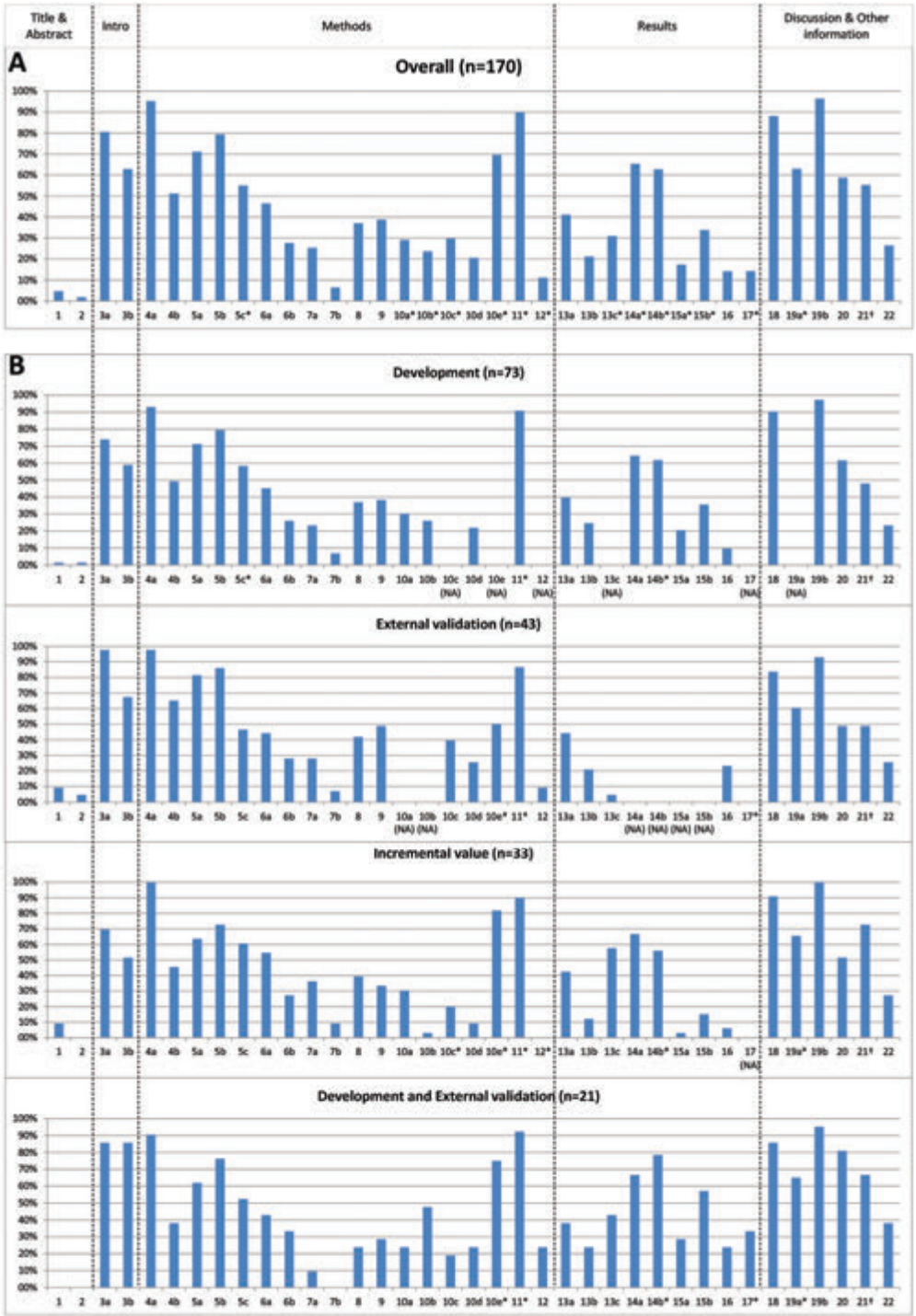
| Complete reporting for >75% of the models | | | Complete reporting for <25% of the models | | |
|---|---|---|---|---|---|
| **TRIPOD items** | | **%** | **TRIPOD items** | | **%** |
| **19b** | Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence. | 96 | **10b** | Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation. | 24 |
| **4a** | Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable. | 95 | **10d** | Specify all measures used to assess model performance and, if relevant, to compare multiple models. | 21 |
| **11** | Provide details on how risk groups were created, if done. | 90 | **13b** | Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome. | 21 |
| **18** | Discuss any limitations of the study (such as non-representative sample, few events per predictor, missing data). | 88 | **15a** | Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point). | 17 |
| **3a** | Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models | 81 | **16** | Report performance measures (with CIs) for the prediction model. | 14 |
| **5b** | Describe eligibility criteria for participants. | 79 | **17** | If done, report the results from any model updating (i.e., model specification, model performance). | 14 |

| | | | |
|---|---|---|---|
| | **12** | For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors. | 11 |
| | **7b** | Report any actions to blind assessment of predictors for the outcome and other predictors. | 6 |
| | **1** | Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted. | 5 |
| | **2** | Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions. | 2 |

Completeness of reporting of individual TRIPOD items is presented in Figure 3 and Additional file 3 over all 170 models, and per type of prediction model study. The most notable findings for each section of the TRIPOD statement (title and abstract, introduction, methods, results, discussion, and other information) are described below.

**Figure 3 (right page). Reporting of the items of the TRIPOD statement overall (A), and per type of prediction model study (B) (see Box 1 for list of items of the TRIPOD statement)**

**NA: not applicable (not all items of the TRIPOD statement are relevant to all types of prediction model studies) Percentages are based on number of models for which an item was applicable (and thus should have been reported).**
***Where this number deviates from the total number of models, this is indicated. This concerns the following items (N=number of models for which the item was applicable): Overall: 5c (N=169), 10a (N=127), 10b (N=127), 10c (N=84), 10e (N=23), 11 (N=70), 12 (N=81), 13c (N=97), 14a (N=127), 14b (N=94), 15a (N=127), 15b (N=127), 17 (N=7), 19a (N=92) Development: 5c (N=72), 11 (N=22), 14b (N=55); External validation: 10e (N=8), 11 (N=15), 17 (N=4); Incremental value: 10c (N=20), 10e (N=11), 11 (N=20), 12 (N=17), 14b (N=25), 19a (N=29); Development and external validation: 10e (N=4), 11 (N=13), 14b (N=14), 17 (N=3), 19a (N=20).**
**†Item 21 "Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets": the number of models for which this item was applicable is unknown. It probably was applicable to all models that reported this item. Instead of presenting a percentage of 100, we based the percentage on the total number of models.**


Title & Abstract
Intro
Methods
Results
Discussion & Other information
A
Overall (n=170)
B
Development (n=73)
External validation (n=43)
Incremental value (n=33)
Development and External validation (n=21)

*Title and abstract (items 1 and 2)*
According to the TRIPOD statement, an informative title contains (synonyms for) the term *risk prediction model*, the *type of prediction model* study (i.e. development, external validation, incremental value, or combination), the *target population,* and *outcome to be predicted*. Eight of the 170 models (5%) addressed all four elements. The description of the type of prediction model study was the least reported element (12%). Complete reporting of abstracts required information for 12 elements. Three of the models (2%) fulfilled all the requirements.

*Introduction (item 3)*
For 81% of the models complete information about background and rationale was provided (item 3a) and in 63% reporting of study objectives (item 3b), including a specification of the type of prediction model study, was considered complete.

*Methods (items 4 – 12)*
Source of data (item 4a; 95% reported) and eligibility criteria (item 5b; 79%) were among the best reported items for all four types of prediction model studies. Actions to blind assessment of (non-objective) outcomes (item 6b; 28%) and predictors (item 7b; 7%) were less well reported. Detailed predictor definitions (item 7a) were provided for 25% of the models. Also information about how missing data were handled (item 9) was incomplete for the majority of models (reported in 39%). Most aspects of statistical analysis were inadequately reported as well. How predictors were handled (item 10a) was described in 29% of the models. Model building procedures (item 10b) were specified in 24% overall, and particularly poor in incremental value reports (3%). Few studies (21%) described both discrimination and calibration as measures of model performance ( item 10d).

*Results (items 13 – 17)*
Characteristics of participants (item 13b, complete reporting in 21%) were often reported without information regarding missing data for predictors and outcome. Two (5%) of the external validations presented demographics, distribution of predictors, and outcomes alongside those of the original development study (item 13c) and in combined reports of development and external validation this was done in 43%. The final model was presented in full (item 15a) in 17% of the models. For many models the intercept (or the cumulative baseline hazard (or baseline survival) for at least one time point in the case of survival models) was not provided. A small number of models provided information on both discrimination and calibration when reporting model

performance (item 16; 14%). Discrimination was more frequently reported (79%) than calibration (29%).

*Discussion (items 18 – 20)*
An overall interpretation of the results (item 19b) was given for almost all included models of all types of prediction model studies (97%). The potential for clinical use and implications for future research (item 20) were discussed in 59% of the models.

*Other information (items 21 and 22)*
Information about the availability of supplementary resources (item 21) was provided in 55% of the models. Complete information regarding funding (item 22) was reported in 27%.



## Discussion

Complete and accurate reporting of prediction model studies is required to critically appraise, externally validate, evaluate their impact, and eventually use prediction models in clinical practice. Our study shows that, regardless of the type of prediction model study and whether diagnostic or prognostic, more than half of the items deemed essential to report in prediction model publications according to the TRIPOD statement were not completely reported.

Highly problematic TRIPOD items in terms of reporting were items regarding title and abstract. These items, for which complete reporting requires information on multiple elements, were adequately reported for less than 10% of the models. In addition, details of study methods, especially blinding of outcome and predictor assessments, were provided for only a minority of reported models. Furthermore, information on follow-up, predictor definitions, model building procedures and handling of missing data were often lacking. Notable findings regarding the reporting of study results were that in over 70% of the included models the final model was not presented in enough detail to make predictions for new patients, and that the reporting of model performance was often incomplete. Items of the TRIPOD statement that were generally well reported addressed the source of data and eligibility criteria, risk groups (if applicable), study limitations, and overall interpretation of results.

### Comparison with other studies

Our main finding of inadequate reporting in the majority of publications within 37 clinical domains is comparable to the findings of systematic reviews of prediction model studies performed in general medicine or specific clinical domains.[6-11]

Inadequate reporting is considered to be a form of research waste.[18,19] Therefore, for many study types reporting guidelines were published in the last 20 years, such as the CONSORT (Consolidated Standards of Reporting Trials) statement in 1996 (updates in 2001 and 2010), the STARD (Standards for Reporting of Diagnostic Accuracy) statement in 2003 (update in 2015), and REMARK (Reporting recommendations for tumour marker prognostic studies) in 2005.[20-24] Completeness of reporting before the introduction of these reporting guidelines was similar to our result of 44% adherence. Moher and colleagues (2001) evaluated 97 reports of randomized trials before the introduction of CONSORT and found adequate reporting for just over half of the items (58%).[25] In a systematic review of 16 studies evaluating the adherence to STARD, overall, 51% of items were adequately reported.[26] For six included studies with quantitative data before publication of STARD a range of 44% to 61% adherence was reported. An assessment of the reporting of prognostic studies of tumour markers was done shortly after the introduction of REMARK.[27,28] Ten (out of 20) items were evaluated, and, overall, articles adhered to 53% of these.

**Strengths and limitations of this study**

With this literature review we cover a broad literature base by including three major types of prediction model studies, both prognostic and diagnostic, across 37 clinical domains. Despite the use of a validated search strategy, we may have missed publications on prediction models. It is likely that the completeness of reporting of prediction models in these studies would have been worse. Furthermore, we selected studies from high impact journals. Therefore, our results on the completeness of reporting might be an optimistic representation of the reporting of prediction model studies in general.

In accordance with the TRIPOD statement, we included prediction models based on regression modelling approaches.[12,13] Although most TRIPOD items would apply, transparent reporting of prediction models using non-regression modelling techniques may require additional details, especially regarding model building procedures, and specific guidance might be desirable.

We were strict in scoring adherence by requiring complete information on all elements of a TRIPOD item, e.g. complete reporting of model performance required the provision of both discrimination and calibration measures. This is in line with the nature of TRIPOD as having essential items needed to appraise and utilize a prediction model. However, authors might have good reasons not to provide specific details regarding an item. For example, if they believe that their model should not be validated or used

in clinical practice, they may have decided not to present the coefficients of the full model. In the current study we would have scored TRIPOD item 15a as "incompletely reported". Although strict scoring potentially leads to poorer adherence results, it is needed for reasons of consistency.

We used two different denominators in our analyses, the number of publications (n=146) and the number of models (n=170). It implies that in the "model" analysis a number of publications were included multiple times. It is likely that results from the same publication although based on the reporting of different models are correlated. Given the descriptive nature of our analysis, we did not adjust for such a possible correlation.



We present results from studies that were published four years ago, nevertheless we expect these findings to be still applicable and relevant to current publications of prediction models. From evaluations of other reporting guidelines, like CONSORT and STARD, we know that it takes time to demonstrate the impact of a reporting guideline on completeness of reporting and changes over several years might be small.[25,26,28-33] To our opinion, therefore, it is too early for a before-after comparison at this moment, and the focus should first be on optimal implementation of TRIPOD.

**Implications for practice and areas for future research**

Inadequate reporting impedes the use of all available evidence regarding a prediction model. First, as title and abstract were among the least well reported items, identifying publications of prediction model studies might be challenging. In addition, we found the reporting of model development often insufficiently detailed, which makes external validation almost impossible. As a consequence, a new model might be developed, rather than making use of an existing model. Also, without model specifications it is impossible to use the model in clinical practice. Finally, inadequate reporting hinders critical appraisal and, by that, the possibility of methodological investigation of sources of variation and bias in prediction model studies.

Experiences from other research areas indicate that the improvement in reporting after the introduction of a guideline is often slow and might be subtle.[25,26,28-33] Improving the completeness of reporting of prediction models is probably even more challenging, as it is a relatively young, less well known research field, with methodology in development and not yet strongly embedded in education. Moreover, the multivariable nature of prediction model studies and their focus on absolute probabilities rather than on comparative measures require the reporting of many

details on methods and results. It should also be taken into account that practical issues, like word limits or journal requirements, could act as barriers for complete reporting.

The introduction of the TRIPOD statement was the first step in improving the reporting of prediction model studies. However, more activities should be undertaken to enhance the implementation of the TRIPOD statement. Active implementation involves a collaborative effort of developers of a reporting guideline and other stakeholders within the academic community, like journal editors and educational institutions. Apart from raising awareness and providing training, possible post-publication activities that are recommended are encouraging guideline endorsement, asking for feedback, and evaluating the impact of the reporting guideline.[34]

By highlighting the flaws in the reporting of prediction model studies, our results enable a targeted implementation strategy for the TRIPOD statement. Possible future activities are the development of educational materials and training regarding specific aspects of the reporting of prediction model studies. The examples of both adequate and suboptimal reporting within our dataset can be used in the training of different stakeholders. An initiative that already has been started by the TRIPOD Group is the development of specific guidance on informative reporting of prediction model studies in abstracts.[35] Furthermore, as TRIPOD is periodically being reappraised and will be updated if necessary, our study will provide useful input for modifications of specific TRIPOD items, related to either content, phrasing or more detailed explanation.[12] Finally, our study will serve as a baseline measurement for future studies evaluating the impact of the introduction the TRIPOD statement.

## Conclusion

Prediction models are poorly reported: more than half of the items that are considered essential for transparent reporting of a prediction model were not or were inadequately reported, especially with regard to details of the title, abstract, blinding, model building procedures, the final model, and model performance. The results of this study can be used to further develop and refine the implementation and increase the impact of the TRIPOD statement.

## Acknowledgement

We thank René Spijker for performing the search for this comprehensive literature survey, and Daan Michels for his assistance with data extraction.

## References


1. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med.* 2013;10(2):e1001381.
2. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ.* 2009;338:b605.
3. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart.* 2012;98(9):691-698.
4. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart.* 2012;98(9):683-690.
5. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009;338:b604.
6. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med.* 2011;9:103.
7. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med.* 2012;9(5):1-12.
8. Collins GS, Omar O, Shanyinde M, Yu LM. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. *J Clin Epidemiol.* 2013;66(3):268-277.
9. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol.* 2014;14:40.
10. Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ.* 2016;353:i2416.
11. Wen Z, Guo Y, Xu B, Xiao K, Peng T, Peng M. Developing Risk Prediction Models for Postoperative Pancreatic Fistula: a Systematic Review of Methodology and Reporting Quality. *Indian J Surg.* 2016;78(2):136-143.
12. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55-63.
13. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1-73.
14. 2012 Journal Citation Reports® In. Science Edition ed: Clarivate Analytics 2017.
15. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc.* 2001;8(4):391-397.

16. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338:b375.
17. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med*. 2013;10(2):e1001380.
18. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374(9683):86-89.
19. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267-276.
20. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.
21. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152(11):726-732.
22. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
23. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.
24. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer*. 2005;93(4):387-391.
25. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA*. 2001;285(15):1992-1995.
26. Korevaar DA, van Enst WA, Spijker R, Bossuyt PM, Hooft L. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evid Based Med*. 2014;19(2):47-54.
27. Mallett S, Timmer A, Sauerbrei W, Altman DG. Reporting of prognostic studies of tumour markers: a review of published articles in relation to REMARK guidelines. *Br J Cancer*. 2010;102(1):173-180.
28. Sekula P, Mallett S, Altman DG, Sauerbrei W. Did the reporting of prognostic studies of tumour markers improve since the introduction of REMARK guideline? A comparison of reporting in published articles. *PLoS One*. 2017;12(6):e0178531.
29. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ*. 2010;340:c723.

30. Korevaar DA, Wang J, van Enst WA, Leeflang MM, Hooft L, et al. Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. *Radiology*. 2015;274(3):781-789.
31. Smidt N, Rutjes AW, van der Windt DA, Ostelo RW, Bossuyt PM, et al. The quality of diagnostic accuracy studies since the STARD statement: has it improved? *Neurology*. 2006;67(5):792-797.
32. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev*. 2012;11:Mr000030.
33. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. *Lancet*. 2005;365(9465):1159-1162.
34. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7(2):e1000217.
35. Heus P, Hooft L, Reitsma JB, Scholten RJPM, Altman DG, et al. Reporting of clinical prediction model studies in journal and conference abstracts: TRIPOD for Abstracts. Paper presented at: 24th Cochrane Colloquium; 2016 23-27 Oct2016; Seoul, Korea.

## Additional Files

### Additional File 1 - Journal selection

**Ten journals with the highest Journal Impact Factor within each of 37 categories (clinical domains) (2012 Journal Citation Reports® [Clarivate Analytics, 2017]) that were selected. Full journal titles indicated with an * were included in more than one category.**

| Category (clinical domain) | Full journal title | Journal Impact Factor |
|---|---|---|
| Allergy | Journal of Allergy and Clinical Immunology* | 12.047 |
| | Allergy | 5.883 |
| | Clinical Reviews in Allergy & Immunology | 5.590 |
| | Clinical and Experimental Allergy | 4.789 |
| | Annals of Allergy Asthma & Immunology | 3.449 |
| | Current Opinion In Allergy and Clinical Immunology | 3.398 |
| | Pediatric Allergy and Immunology* | 3.376 |
| | Contact Dermatitis* | 2.925 |
| | Current Allergy and Asthma Reports | 2.746 |
| | Allergy Asthma & Immunology Research | 2.653 |
| Anesthesiology | Pain | 5.644 |
| | Anesthesiology | 5.163 |
| | British Journal of Anaesthesia | 4.237 |
| | Anaesthesia | 3.486 |
| | Regional Anesthesia and Pain Medicine | 3.464 |
| | Anesthesia and Analgesia | 3.300 |
| | European Journal of Pain | 3.067 |
| | Minerva Anestesiologica* | 2.818 |
| | European Journal of Anaesthesiology | 2.792 |
| | Pain Practice | 2.605 |
| Cardiac and cardiovasuclar systems | Circulation* | 15.202 |
| | European Heart Journal | 14.097 |
| | Journal of the American College of Cardiology | 14.086 |
| | Circulation Research* | 11.861 |
| | Nature Reviews Cardiology | 10.400 |
| | Circulation-Cardiovascular Genetics | 6.728 |
| | Circulation-Heart Failure | 6.684 |
| | Jacc-Cardiovascular Interventions | 6.552 |
| | Circulation-Cardiovascular Interventions | 6.543 |
| | Jacc-Cardiovascular Imaging* | 6.164 |

| | | |
|---|---|---|
| Clinical neurology | Lancet Neurology | 23.917 |
| | Nature Reviews Neurology | 15.518 |
| | Alzheimers & Dementia | 14.483 |
| | Annals of Neurology | 11.193 |
| | Brain | 9.915 |
| | Acta Neuropathologica | 9.734 |
| | Sleep Medicine Reviews | 8.681 |
| | Neurology | 8.249 |
| | Archives of Neurology | 7.685 |
| | Neuro-Oncology | 6.180 |
| Critical care medicine | American Journal of Respiratory and Critical Care Medicine* | 11.041 |
| | Critical Care Medicine | 6.124 |
| | Chest* | 5.854 |
| | Intensive Care Medicine | 5.258 |
| | Critical Care | 4.718 |
| | Journal of Neurotrauma | 4.295 |
| | Resuscitation* | 4.104 |
| | Neurocritical Care | 3.038 |
| | Current Opinion In Critical Care | 2.967 |
| | Minerva Anestesiologica* | 2.818 |
| Dentistry. Oral surgery & medicine | Periodontology 2000 | 4.012 |
| | Journal of Dental Research | 3.826 |
| | Clinical Implant Dentistry and Related Research | 3.821 |
| | Dental Materials | 3.773 |
| | Journal of Clinical Periodontology | 3.688 |
| | Clinical Oral Implants Research | 3.433 |
| | Journal of Dentistry | 3.200 |
| | Journal of Endodontics | 2.929 |
| | International Journal of Oral Science | 2.719 |
| | British Journal of Oral & Maxillofacial Surgery | 2.717 |
| Dermatology | Journal of Investigative Dermatology | 6.193 |
| | Pigment Cell & Melanoma Research | 5.839 |
| | Journal of the American Academy of Dermatology | 4.906 |
| | Archives of Dermatology | 4.792 |
| | British Journal of Dermatology | 3.759 |
| | Experimental Dermatology | 3.578 |
| | Journal of Dermatological Science | 3.520 |
| | Acta Dermato-Venereologica | 3.487 |
| | Contact Dermatitis* | 2.925 |
| | Skin Pharmacology and Physiology | 2.885 |

| | | |
|---|---|---|
| Emergency medicine | Annals of Emergency Medicine | 4.285 |
| | Resuscitation* | 4.104 |
| | Emergencias | 2.578 |
| | Journal of Trauma-Injury Infection and Critical Care | 2.348 |
| | Injury-International Journal of the Care of the Injured | 2.174 |
| | Prehospital Emergency Care | 1.859 |
| | Academic Emergency Medicine | 1.757 |
| | American Journal of Emergency Medicine | 1.704 |
| | Scandinavian Journal of Trauma Resuscitation & Emergency Medicine | 1.680 |
| | Emergency Medicine Journal | 1.645 |
| Endocrinology & Metabolism | Endocrine Reviews | 14.873 |
| | Cell Metabolism | 14.619 |
| | Nature Reviews Endocrinology | 11.025 |
| | Trends In Endocrinology and Metabolism | 8.901 |
| | Frontiers In Neuroendocrinology | 7.985 |
| | Diabetes | 7.895 |
| | Diabetes Care | 7.735 |
| | Journal of Mammary Gland Biology and Neoplasia | 7.524 |
| | Journal of Pineal Research | 7.304 |
| | Antioxidants & Redox Signaling | 7.189 |
| Gastroenterology & Hepatology | Gastroenterology | 12.821 |
| | Hepatology | 12.003 |
| | Gut | 10.732 |
| | Nature Reviews Gastroenterology & Hepatology | 10.426 |
| | Journal of Hepatology | 9.858 |
| | Seminars In Liver Disease | 8.274 |
| | American Journal of Gastroenterology | 7.553 |
| | Clinical Gastroenterology and Hepatology | 6.648 |
| | Endoscopy* | 5.735 |
| | Gastrointestinal Endoscopy | 5.210 |
| Geriatrics & Gerontology | Neurobiology of Aging | 6.166 |
| | Ageing Research Reviews | 5.953 |
| | Aging Cell | 5.705 |
| | Journal of the American Medical Directors Association | 5.302 |
| | Frontiers In Aging Neuroscience | 5.224 |
| | Journals of Gerontology Series A-Biological Sciences and Medical Sciences | 4.314 |
| | American Journal of Geriatric Psychiatry | 4.131 |
| | Age | 4.084 |
| | Journal of the American Geriatrics Society | 3.978 |
| | Experimental Gerontology | 3.911 |

| | | |
|---|---|---|
| Hematology | Circulation Research* | 11.861 |
| | Leukemia* | 10.164 |
| | Blood | 9.060 |
| | Stem Cells | 7.701 |
| | Arteriosclerosis Thrombosis and Vascular Biology* | 6.338 |
| | Thrombosis and Haemostasis* | 6.094 |
| | Journal of Thrombosis and Haemostasis* | 6.081 |
| | Blood Reviews | 6.000 |
| | Haematologica-the Hematology Journal | 5.935 |
| | Journal of Cerebral Blood Flow and Metabolism | 5.398 |
| Immunology | Annual Review of Immunology | 36.556 |
| | Nature Reviews Immunology | 33.129 |
| | Nature Immunology | 26.199 |
| | Immunity | 19.795 |
| | Journal of Experimental Medicine | 13.214 |
| | Immunological Reviews | 12.155 |
| | Journal of Allergy and Clinical Immunology* | 12.047 |
| | Trends In Immunology | 9.486 |
| | Clinical Infectious Diseases* | 9.374 |
| | Current Opinion In Immunology | 8.771 |
| Infectious diseases | Lancet Infectious Diseases | 19.966 |
| | Clinical Infectious Diseases* | 9.374 |
| | Aids | 6.407 |
| | Emerging Infectious Diseases | 5.993 |
| | Journal of Infectious Diseases | 5.848 |
| | Eurosurveillance | 5.491 |
| | Journal of Antimicrobial Chemotherapy | 5.338 |
| | Current Opinion In Infectious Diseases | 4.870 |
| | Current Opinion In Hiv and Aids | 4.704 |
| | Jaids-Journal of Acquired Immune Deficiency Syndromes | 4.653 |
| Integrative & complementary medicine | Alternative Medicine Review | 4.857 |
| | Phytomedicine | 2.972 |
| | Journal of Ethnopharmacology | 2.755 |
| | Integrative Cancer therapies | 2.354 |
| | American Journal of Chinese Medicine | 2.281 |
| | Complementary therapies In Medicine | 2.093 |
| | Bmc Complementary and Alternative Medicine | 2.082 |
| | Evidence-Based Complementary and Alternative Medicine | 1.722 |
| | Journal of Manipulative and Physiological therapeutics | 1.647 |
| | Journal of Alternative and Complementary Medicine | 1.464 |

| | | |
|---|---|---|
| Medical laboratory technology | Clinical Chemistry | 7.149 |
| | Critical Reviews In Clinical Laboratory Sciences | 3.783 |
| | Advances In Clinical Chemistry | 3.674 |
| | Translational Research | 3.490 |
| | Clinical Chemistry and Laboratory Medicine | 3.009 |
| | Clinica Chimica Acta | 2.850 |
| | Archives of Pathology & Laboratory Medicine | 2.781 |
| | Clinical Biochemistry | 2.450 |
| | Therapeutic Drug Monitoring | 2.234 |
| | Cytometry Part B-Clinical Cytometry | 2.231 |
| Medicine. general & internal | New England Journal of Medicine | 51.658 |
| | Lancet | 39.060 |
| | Jama-Journal of the American Medical Association | 29.978 |
| | British Medical Journal | 17.215 |
| | Plos Medicine | 15.253 |
| | Annals of Internal Medicine | 13.976 |
| | Archives of Internal Medicine | 10.579 |
| | Bmc Medicine | 6.679 |
| | Canadian Medical Association Journal | 6.465 |
| | Journal of Internal Medicine | 6.455 |
| Obstetrics & Gynecology | Human Reproduction Update* | 8.847 |
| | Obstetrics and Gynecology | 4.798 |
| | Human Reproduction* | 4.670 |
| | Fertility and Sterility* | 4.174 |
| | Gynecologic Oncology | 3.929 |
| | American Journal of Obstetrics and Gynecology | 3.877 |
| | Bjog-An International Journal of Obstetrics and Gynaecology | 3.760 |
| | Ultrasound In Obstetrics & Gynecology | 3.557 |
| | Seminars In Reproductive Medicine* | 3.211 |
| | Menopause-the Journal of the North American Menopause Society | 3.163 |
| Oncology | Ca-A Cancer Journal For Clinicians | 153.459 |
| | Nature Reviews Cancer | 35.000 |
| | Lancet Oncology | 25.117 |
| | Cancer Cell | 24.755 |
| | Journal of Clinical Oncology | 18.038 |
| | Nature Reviews Clinical Oncology | 15.031 |
| | Jnci-Journal of the National Cancer Institute | 14.336 |
| | Leukemia* | 10.164 |
| | Cancer Discovery | 10.143 |
| | Biochimica Et Biophysica Acta-Reviews On Cancer | 9.033 |

| | | |
|---|---|---|
| Ophthalmology | Progress In Retinal and Eye Research | 9.439 |
| | Ophthalmology | 5.563 |
| | Archives of Ophthalmology | 3.826 |
| | American Journal of Ophthalmology | 3.631 |
| | Investigative Ophthalmology & Visual Science | 3.441 |
| | Experimental Eye Research | 3.026 |
| | Survey of Ophthalmology | 2.859 |
| | Retina-the Journal of Retinal and Vitreous Diseases | 2.825 |
| | British Journal of Ophthalmology | 2.725 |
| | Ocular Surface | 2.643 |
| Orthopedics | American Journal of Sports Medicine* | 4.439 |
| | Osteoarthritis and Cartilage* | 4.262 |
| | Journal of Bone and Joint Surgery-American Volume | 3.234 |
| | Spine Journal | 3.220 |
| | Arthroscopy-the Journal of Arthroscopic and Related Surgery | 3.103 |
| | Journal of Orthopaedic & Sports Physical therapy* | 2.947 |
| | Journal of Orthopaedic Research | 2.875 |
| | Clinical Orthopaedics and Related Research | 2.787 |
| | Physical therapy* | 2.778 |
| | Acta Orthopaedica | 2.736 |
| Otorhinolaryngology | Ear and Hearing | 3.262 |
| | Jaro-Journal of the Association For Research In Otolaryngology | 2.952 |
| | Head and Neck-Journal For the Sciences and Specialties of the Head and Neck | 2.833 |
| | Hearing Research | 2.537 |
| | Audiology and Neuro-Otology | 2.318 |
| | Otology & Neurotology | 2.014 |
| | Laryngoscope | 1.979 |
| | Dysphagia | 1.938 |
| | Clinical Otolaryngology | 1.869 |
| | Archives of Otolaryngology-Head & Neck Surgery | 1.779 |

| | | |
|---|---|---|
| Pediatrics | Journal of the American Academy of Child and Adolescent Psychiatry* | 6.970 |
| | Pediatrics | 5.119 |
| | Archives of Pediatrics & Adolescent Medicine | 4.282 |
| | Journal of Pediatrics | 4.035 |
| | European Child & Adolescent Psychiatry | 3.699 |
| | Pediatric Infectious Disease Journal | 3.569 |
| | Seminars In Fetal & Neonatal Medicine | 3.505 |
| | Archives of Disease In Childhood-Fetal and Neonatal Edition | 3.451 |
| | Pediatric Allergy and Immunology* | 3.376 |
| | Archives of Disease In Childhood | 3.051 |
| Peripheral vascular disease | Circulation* | 15.202 |
| | Circulation Research* | 11.861 |
| | Hypertension | 6.873 |
| | Arteriosclerosis Thrombosis and Vascular Biology* | 6.338 |
| | Stroke | 6.158 |
| | Thrombosis and Haemostasis* | 6.094 |
| | Journal of Thrombosis and Haemostasis* | 6.081 |
| | Current Opinion In Lipidology | 5.839 |
| | Atherosclerosis Supplements | 4.333 |
| | Seminars In Thrombosis and Hemostasis | 4.216 |
| Primary health care | Annals of Family Medicine | 4.613 |
| | Primary Care Respiratory Journal | 2.191 |
| | British Journal of General Practice | 2.034 |
| | Scandinavian Journal of Primary Health Care | 1.905 |
| | Family Practice | 1.828 |
| | Canadian Family Physician | 1.808 |
| | Journal of the American Board of Family Medicine | 1.758 |
| | American Family Physician | 1.611 |
| | Bmc Family Practice | 1.609 |
| | Primary Care Diabetes | 1.609 |
| Psychiatry | Molecular Psychiatry | 14.897 |
| | American Journal of Psychiatry | 14.721 |
| | Archives of General Psychiatry | 13.772 |
| | Biological Psychiatry | 9.247 |
| | World Psychiatry | 8.974 |
| | Neuropsychopharmacology | 8.678 |
| | Schizophrenia Bulletin | 8.486 |
| | Psychotherapy and Psychosomatics | 7.230 |
| | Journal of the American Academy of Child and Adolescent Psychiatry* | 6.970 |
| | British Journal of Psychiatry | 6.606 |

| | | |
|---|---|---|
| Public. Environmental and Occupational health | Epidemiologic Reviews | 9.269 |
| | Environmental Health Perspectives | 7.260 |
| | International Journal of Epidemiology | 6.982 |
| | Who Technical Report Series | 6.100 |
| | Epidemiology | 5.738 |
| | Journal of Clinical Epidemiology | 5.332 |
| | Bulletin of the World Health Organization | 5.250 |
| | European Journal of Epidemiology | 5.118 |
| | American Journal of Epidemiology | 4.780 |
| | Cancer Epidemiology Biomarkers & Prevention | 4.559 |
| Radiology. Nuclear medicine and Medical imaging | Human Brain Mapping | 6.878 |
| | Radiology | 6.339 |
| | Neuroimage | 6.252 |
| | Jacc-Cardiovascular Imaging* | 6.164 |
| | Circulation-Cardiovascular Imaging | 5.795 |
| | Journal of Nuclear Medicine | 5.774 |
| | Investigative Radiology | 5.460 |
| | European Journal of Nuclear Medicine and Molecular Imaging | 5.114 |
| | International Journal of Radiation Oncology Biology Physics | 4.524 |
| | Radiotherapy and Oncology | 4.520 |
| Rehabilitation | Journal of Head Trauma Rehabilitation | 4.443 |
| | Neurorehabilitation and Neural Repair | 4.278 |
| | Ieee Transactions On Neural Systems and Rehabilitation Engineering | 3.255 |
| | Journal of Orthopaedic & Sports Physical therapy* | 2.947 |
| | Physical therapy* | 2.778 |
| | Supportive Care In Cancer | 2.649 |
| | Journal of Neuroengineering and Rehabilitation | 2.567 |
| | American Journal of Speech-Language Pathology | 2.448 |
| | Archives of Physical Medicine and Rehabilitation | 2.358 |
| | Journal of Physiotherapy | 2.255 |
| Reproductive biology | Human Reproduction Update* | 8.847 |
| | Human Reproduction* | 4.670 |
| | Molecular Human Reproduction | 4.542 |
| | Fertility and Sterility* | 4.174 |
| | Biology of Reproduction | 4.027 |
| | Reproduction | 3.555 |
| | American Journal of Reproductive Immunology | 3.317 |
| | Seminars In Reproductive Medicine* | 3.211 |
| | Reproductive Toxicology | 3.141 |
| | Placenta | 3.117 |

| | | |
|---|---|---|
| Respiratory system | American Journal of Respiratory and Critical Care Medicine* | 11.041 |
| | Thorax | 8.376 |
| | European Respiratory Journal | 6.355 |
| | Chest* | 5.854 |
| | Journal of Heart and Lung Transplantation* | 5.112 |
| | Journal of Thoracic Oncology | 4.473 |
| | American Journal of Respiratory Cell and Molecular Biology | 4.148 |
| | Respiratory Research | 3.642 |
| | Journal of Thoracic and Cardiovascular Surgery | 3.526 |
| | American Journal of Physiology-Lung Cellular and Molecular Physiology | 3.523 |
| Rheumatology | Nature Reviews Rheumatology | 9.745 |
| | Annals of the Rheumatic Diseases | 9.111 |
| | Arthritis and Rheumatism | 7.477 |
| | Current Opinion In Rheumatology | 5.191 |
| | Arthritis Research & therapy | 4.302 |
| | Osteoarthritis and Cartilage* | 4.262 |
| | Rheumatology | 4.212 |
| | Seminars In Arthritis and Rheumatism | 3.806 |
| | Arthritis Care & Research | 3.731 |
| | Best Practice & Research In Clinical Rheumatology | 3.550 |
| Sport sciences | Exercise Immunology Review | 7.053 |
| | Exercise and Sport Sciences Reviews | 5.283 |
| | Sports Medicine | 5.237 |
| | Medicine and Science In Sports and Exercise | 4.475 |
| | American Journal of Sports Medicine* | 4.439 |
| | British Journal of Sports Medicine | 3.668 |
| | Journal of Applied Physiology | 3.484 |
| | Scandinavian Journal of Medicine & Science In Sports | 3.214 |
| | Journal of Orthopaedic & Sports Physical therapy* | 2.947 |
| | Journal of Science and Medicine In Sport | 2.899 |
| Surgery | Annals of Surgery | 6.329 |
| | American Journal of Transplantation* | 6.192 |
| | Endoscopy* | 5.735 |
| | Journal of Neurology Neurosurgery and Psychiatry | 4.924 |
| | American Journal of Surgical Pathology | 4.868 |
| | British Journal of Surgery | 4.839 |
| | Journal of the American College of Surgeons | 4.500 |
| | Surgery For Obesity and Related Diseases | 4.121 |
| | Annals of Surgical Oncology | 4.120 |
| | Archives of Surgery | 4.100 |

| | | |
|---|---|---|
| Transplantation | American Journal of Transplantation* | 6.192 |
| | Journal of Heart and Lung Transplantation* | 5.112 |
| | Stem Cells and Development | 4.670 |
| | Cell Transplantation | 4.422 |
| | Liver Transplantation | 3.944 |
| | Biology of Blood and Marrow Transplantation | 3.940 |
| | Transplantation | 3.781 |
| | Bone Marrow Transplantation | 3.541 |
| | Nephrology Dialysis Transplantation | 3.371 |
| | Current Opinion In Organ Transplantation | 3.272 |
| Tropical medicine | Plos Neglected Tropical Diseases | 4.569 |
| | Malaria Journal | 3.400 |
| | Tropical Medicine & International Health | 2.938 |
| | Acta Tropica | 2.787 |
| | American Journal of Tropical Medicine and Hygiene | 2.534 |
| | Transactions of the Royal Society of Tropical Medicine and Hygiene | 1.823 |
| | Memorias Do Instituto Oswaldo Cruz | 1.363 |
| | Annals of Tropical Medicine and Parasitology | 1.313 |
| | Journal of Vector Borne Diseases | 1.041 |
| | Journal of Tropical Pediatrics | 1.006 |
| Urology & Nephrology | European Urology | 10.476 |
| | Journal of the American Society of Nephrology | 8.987 |
| | Nature Reviews Nephrology | 7.943 |
| | Kidney International | 7.916 |
| | American Journal of Kidney Diseases | 5.294 |
| | Clinical Journal of the American Society of Nephrology | 5.068 |
| | Nature Reviews Urology | 4.793 |
| | Current Opinion In Nephrology and Hypertension | 3.964 |
| | Prostate | 3.843 |
| | Journal of Urology | 3.696 |

**Additional File 2 - Search strategy**
**Pubmed search strategy on July 4th 2014**

| | hits |
|---|---|
| ((Validat*[tiab] OR Predict*[ti] OR Rule*[tiab]) OR (Predict*[tiab] AND (Outcome*[tiab] OR Risk*[tiab] OR Model*[tiab])) OR ((History[tiab] OR Variable*[tiab] OR Criteria[tiab] OR Scor*[tiab] OR Characteristic*[tiab] OR Finding*[tiab] OR Factor*[tiab]) AND (Predict*[tiab] OR Model*[tiab] OR Decision*[tiab] OR Identif*[tiab] OR Prognos*[tiab])) OR (Decision*[tiab] AND (Model*[tiab] OR Clinical*[tiab] OR logistic models[mesh])) OR (Prognostic[tiab] AND (History[tiab] OR Variable*[tiab] OR Criteria[tiab] OR Scor*[tiab] OR Characteristic*[tiab] OR Finding*[tiab] OR Factor*[tiab] OR Model*[tiab]))) AND (0091-6749[is] OR 0105-4538[is] OR 1080-0549[is] OR 0954-7894[is] OR 1081-1206[is] OR 1528-4050[is] OR 0905-6157[is] OR 0105-1873[is] OR 1529-7322[is] OR 2092-7355[is] OR 0304-3959[is] OR 0003-3022[is] OR 0007-0912[is] OR 0003-2409[is] OR 1098-7339[is] OR 0003-2999[is] OR 1090-3801[is] OR 0375-9393[is] OR 0265-0215[is] OR 1530-7085[is] OR 0009-7322[is] OR 0195-668X[is] OR 0735-1097[is] OR 0009-7330[is] OR 1759-5002[is] OR 1942-325X[is] OR 1941-3289[is] OR 1936-8798[is] OR 1941-7640[is] OR 1936-878X[is] OR 1474-4422[is] OR 1759-4758[is] OR 1552-5260[is] OR 0364-5134[is] OR 0006-8950[is] OR 0001-6322[is] OR 1087-0792[is] OR 0028-3878[is] OR 0003-9942[is] OR 1522-8517[is] OR 1073-449X[is] OR 0090-3493[is] OR 0012-3692[is] OR 0342-4642[is] OR 1466-609X[is] OR 0897-7151[is] OR 0300-9572[is] OR 1541-6933[is] OR 1070-5295[is] OR 0375-9393[is] OR 0906-6713[is] OR 0022-0345[is] OR 1523-0899[is] OR 0109-5641[is] OR 0303-6979[is] OR 0905-7161[is] OR 0300-5712[is] OR 0099-2399[is] OR 1674-2818[is] OR 0266-4356[is] OR 0022-202X[is] OR 1755-1471[is] OR 0190-9622[is] OR 0003-987X[is] OR 0007-0963[is] OR 0906-6705[is] OR 0923-1811[is] OR 0001-5555[is] OR 0105-1873[is] OR 1660-5527[is] OR 0196-0644[is] OR 0300-9572[is] OR 1137-6821[is] OR 0022-5282[is] OR 0020-1383[is] OR 1090-3127[is] OR 1069-6563[is] OR 0735-6757[is] OR 1757-7241[is] OR 1472-0205[is] OR 0163-769X[is] OR 1550-4131[is] OR 1759-5029[is] OR 1043-2760[is] OR 0091-3022[is] OR 0012-1797[is] OR 0149-5992[is] OR 1083-3021[is] OR 0742-3098[is] OR 1523-0864[is] OR 0016-5085[is] OR 0270-9139[is] OR 0017-5749[is] OR 1759-5045[is] OR 0168-8278[is] OR 0272-8087[is] OR 0002-9270[is] OR 1542-3565[is] OR 0013-726X[is] OR 0016-5107[is] OR 0028-4793[is] OR 0140-6736[is] OR 0098-7484[is] OR 1756-1833[is] OR 1549-1676[is] OR 0003-4819[is] OR 0003-9926[is] OR 1741-7015[is] OR 0820-3946[is] OR 0954-6820[is] OR 0197-4580[is] OR 1568-1637[is] OR 1474-9718[is] OR 1525-8610[is] OR 1663-4365[is] OR 1079-5006[is] OR 1064-7481[is] OR 0161-9152[is] OR 0002-8614[is] OR 0531-5565[is] OR 0009-7330[is] OR 0887-6924[is] OR 0006-4971[is] OR 1066-5099[is] OR 1079-5642[is] OR 0340-6245[is] OR 1538-7933[is] OR 0268-960X[is] OR 0390-6078[is] OR 0271-678X[is] OR 0732-0582[is] OR 1474-1733[is] OR 1529-2908[is] OR 1074-7613[is] OR 0022-1007[is] OR 0105-2896[is] OR 0091-6749[is] OR 1471-4906[is] OR 1058-4838[is] OR 0952-7915[is] OR 1473-3099[is] OR 1058-4838[is] OR 0269-9370[is] OR 1080-6040[is] OR 0022-1899[is] OR 1560-7917[is] OR 0305-7453[is] OR 0951-7375[is] OR 1746-630X[is] OR 1525-4135[is] OR 1089-5159[is] OR 0944-7113[is] OR 0378-8741[is] OR 1534-7354[is] OR 0192-415X[is] OR 0965-2299[is] OR 1472-6882[is] OR 1741-427X[is] OR 0161-4754[is] OR 1075-5535[is] OR 0009-9147[is] | 4871 |

OR 1040-8363[is] OR 0065-2423[is] OR 1931-5244[is] OR 1434-6621[is] OR 0009-8981[is] OR 0003-9985[is] OR 0009-9120[is] OR 0163-4356[is] OR 1552-4949[is] OR 1355-4786[is] OR 0029-7844[is] OR 0268-1161[is] OR 0015-0282[is] OR 0090-8258[is] OR 0002-9378[is] OR 1470-0328[is] OR 0960-7692[is] OR 1526-8004[is] OR 1072-3714[is] OR 0007-9235[is] OR 1474-175X[is] OR 1470-2045[is] OR 1535-6108[is] OR 0732-183X[is] OR 1759-4774[is] OR 0027-8874[is] OR 0887-6924[is] OR 2159-8274[is] OR 1350-9462[is] OR 0161-6420[is] OR 0003-9950[is] OR 0002-9394[is] OR 0146-0404[is] OR 0014-4835[is] OR 0039-6257[is] OR 0275-004X[is] OR 0007-1161[is] OR 1542-0124[is] OR 0363-5465[is] OR 1063-4584[is] OR 0021-9355[is] OR 1529-9430[is] OR 0749-8063[is] OR 0190-6011[is] OR 0736-0266[is] OR 0009-921X[is] OR 0031-9023[is] OR 1745-3674[is] OR 0196-0202[is] OR 1525-3961[is] OR 1043-3074[is] OR 0378-5955[is] OR 1420-3030[is] OR 1531-7129[is] OR 0023-852X[is] OR 0179-051X[is] OR 1749-4478[is] OR 0886-4470[is] OR 0890-8567[is] OR 0031-4005[is] OR 1072-4710[is] OR 0022-3476[is] OR 1018-8827[is] OR 0891-3668[is] OR 1744-165X[is] OR 1359-2998[is] OR 0905-6157[is] OR 0003-9888[is] OR 0009-7322[is] OR 0009-7330[is] OR 0194-911X[is] OR 1079-5642[is] OR 0039-2499[is] OR 0340-6245[is] OR 1538-7933[is] OR 0957-9672[is] OR 1567-5688[is] OR 0094-6176[is] OR 1544-1709[is] OR 1471-4418[is] OR 0960-1643[is] OR 0281-3432[is] OR 0263-2136[is] OR 0008-350X[is] OR 1557-2625[is] OR 0002-838X[is] OR 1471-2296[is] OR 1751-9918[is] OR 1359-4184[is] OR 0002-953X[is] OR 0003-990X[is] OR 0006-3223[is] OR 1723-8617[is] OR 0893-133X[is] OR 0586-7614[is] OR 0033-3190[is] OR 0890-8567[is] OR 0007-1250[is] OR 0193-936X[is] OR 0091-6765[is] OR 0300-5771[is] OR 0512-3054[is] OR 1044-3983[is] OR 0895-4356[is] OR 0042-9686[is] OR 0393-2990[is] OR 0002-9262[is] OR 1055-9965[is] OR 1065-9471[is] OR 0033-8419[is] OR 1053-8119[is] OR 1936-878X[is] OR 1941-9651[is] OR 0161-5505[is] OR 0020-9996[is] OR 1619-7070[is] OR 0360-3016[is] OR 0167-8140[is] OR 0885-9701[is] OR 1545-9683[is] OR 1534-4320[is] OR 0190-6011[is] OR 0031-9023[is] OR 0941-4355[is] OR 1743-0003[is] OR 1058-0360[is] OR 0003-9993[is] OR 1836-9553[is] OR 1355-4786[is] OR 0268-1161[is] OR 1360-9947[is] OR 0015-0282[is] OR 0006-3363[is] OR 1470-1626[is] OR 1046-7408[is] OR 1526-8004[is] OR 0890-6238[is] OR 0143-4004[is] OR 1073-449X[is] OR 0040-6376[is] OR 0903-1936[is] OR 0012-3692[is] OR 1053-2498[is] OR 1556-0864[is] OR 1044-1549[is] OR 1465-993X[is] OR 0022-5223[is] OR 1040-0605[is] OR 1759-4790[is] OR 0003-4967[is] OR 0004-3591[is] OR 1040-8711[is] OR 1478-6354[is] OR 1063-4584[is] OR 1462-0324[is] OR 0049-0172[is] OR 2151-464X[is] OR 1521-6942[is] OR 1077-5552[is] OR 0091-6331[is] OR 0112-1642[is] OR 0195-9131[is] OR 0363-5465[is] OR 0306-3674[is] OR 8750-7587[is] OR 0905-7188[is] OR 0190-6011[is] OR 1440-2440[is] OR 0003-4932[is] OR 1600-6135[is] OR 0013-726X[is] OR 0022-3050[is] OR 0147-5185[is] OR 0007-1323[is] OR 1072-7515[is] OR 1550-7289[is] OR 1068-9265[is] OR 0004-0010[is] OR 1600-6135[is] OR 1053-2498[is] OR 1547-3287[is] OR 0963-6897[is] OR 1527-6465[is] OR 1083-8791[is] OR 0041-1337[is] OR 0268-3369[is] OR 0931-0509[is] OR 1087-2418[is] OR 1935-2735[is] OR 1475-2875[is] OR 1360-2276[is] OR 0001-706X[is] OR 0002-9637[is] OR 0035-9203[is] OR 0074-0276[is] OR 0003-4983[is] OR 0972-9062[is] OR 0142-6338[is] OR 0302-2838[is] OR 1046-6673[is] OR 1759-5061[is] OR 0085-2538[is] OR 0272-6386[is] OR 1555-9041[is] OR 1759-4812[is] OR 1062-4821[is] OR 0270-4137[is] OR 0022-5347[is]) AND (2014/05/01 : 2014/06/01[dp])

**Additional File 3 - Reporting of the items of the TRIPOD statement**

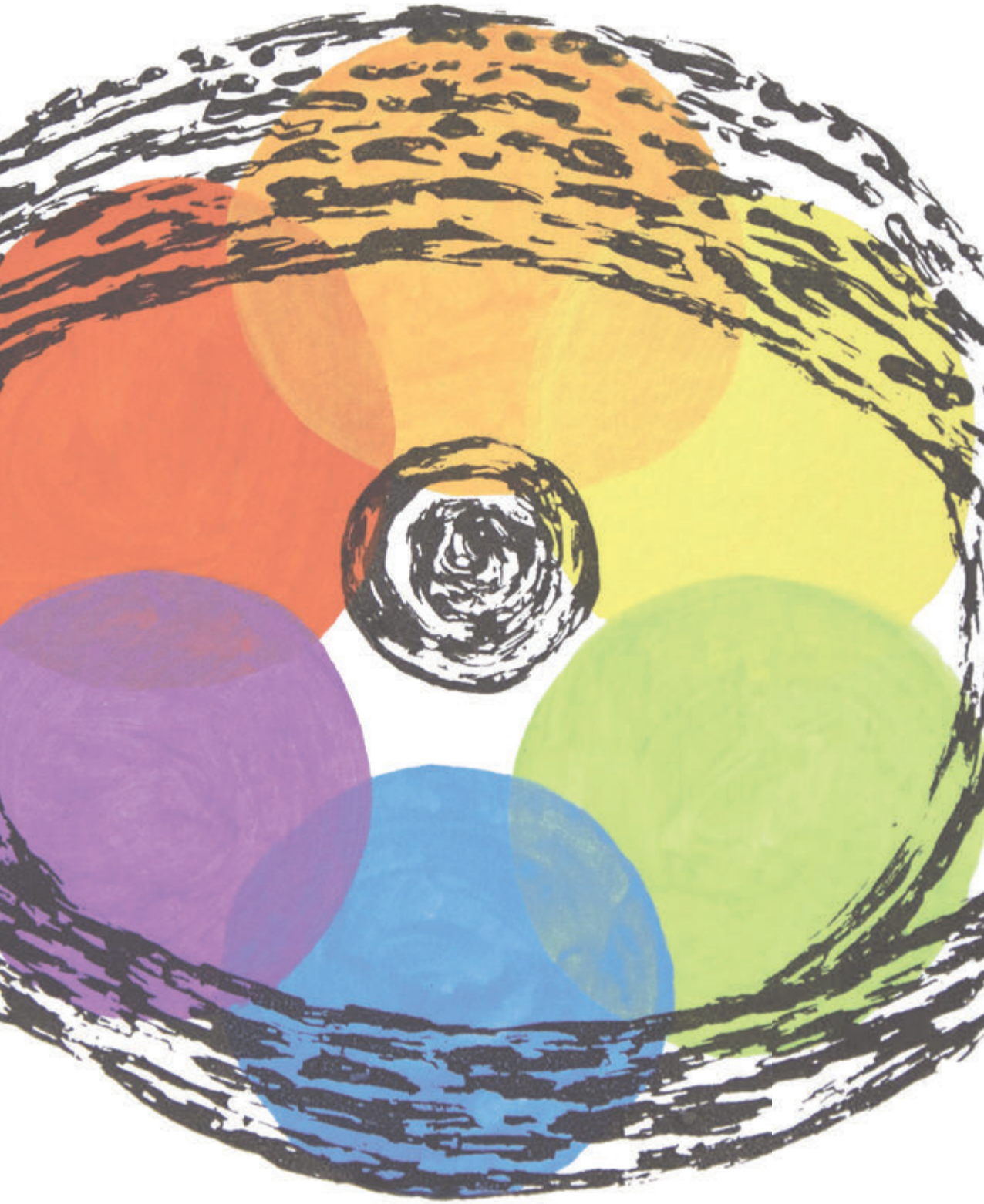
| | Development | External validation | Incremental value | Development and external validation | Overall |
|---|---|---|---|---|---|
| | N=73 | N=43 | N=33 | N=21 | N=170 |
| **Items of the TRIPOD statement** | n (%) | n (%) | n (%) | n (%) | n (%) |
| *Title and abstract* | | | | | |
| 1. **Title**: identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted. | 1 (1) | 4 (9) | 3 (9) | 0 (0) | 8 (5) |
| 2. **Abstract**: provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions. | 7 (10) | 4 (9) | 1 (3) | 1 (5) | 13 (8) |
| *Introduction* | | | | | |
| 3. **Background and objectives**: | | | | | |
| a. Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models. | 54 (74) | 42 (98) | 23 (70) | 18 (86) | 137 (81) |
| b. Specify the objectives, including whether the study describes the development or validation of the model or both. | 43 (59) | 29 (67) | 17 (52) | 18 (86) | 107 (63) |
| *Methods* | | | | | |
| 4. **Source of data**: | | | | | |
| a. Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable. | 68 (93) | 42 (98) | 33 (100) | 19 (91) | 162 (95) |
| b. Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up. | 36 (49) | 28 (65) | 15 (46) | 8 (38) | 87 (51) |

| | | | | | |
|---|---|---|---|---|---|
| **5. Participants**: | | | | | |
| **a.** Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres. | 52 (71) | 35 (81) | 21 (64) | 13 (62) | 121 (71) |
| **b.** Describe eligibility criteria for participants. | 58 (80) | 37 (86) | 24 (73) | 16 (76) | 135 (79) |
| **c.** Give details of treatments received, if relevant. | 42/72 (58)* | 20 (47) | 20 (61) | 11 (52) | 93/169 (55)* |
| **6. Outcome**: | | | | | |
| **a.** Clearly define the outcome that is predicted by the prediction model, including how and when assessed. | 33 (45) | 19 (44) | 18 (55) | 9 (43) | 79 (47) |
| **b.** Report any actions to blind assessment of the outcome to be predicted. | 19 (26) | 12 (28) | 9 (27) | 7 (33) | 47 (28) |
| **7. Predictors**: | | | | | |
| **a.** Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured. | 17 (23) | 12 (28) | 12 (36) | 2 (10) | 43 (25) |
| **b.** Report any actions to blind assessment of predictors for the outcome and other predictors. | 5 (7) | 3 (7) | 3 (9) | 0 (0) | 11 (7) |
| **8. Sample size**: explain how the study size was arrived at. | 27 (37) | 18 (42) | 13 (39) | 5 (24) | 63 (37) |
| **9. Missing data**: Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method. | 28 (38) | 21 (49) | 11 (33) | 6 (29) | 66 (39) |
| **10. Statistical analysis methods**: | | | | | |
| **a.** Describe how predictors were handled in the analyses. | 22 (30) | NA | 10 (30) | 5 (24) | 37/127 (29)* |
| **b.** Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation. | 19 (26) | NA | 1 (3) | 10 (48) | 30/127 (24)* |
| **c.** For validation, describe how the predictions were calculated. | NA | 17 (40) | 4/20 (20)* | 4 (19) | 25/84 (30)* |
| **d.** Specify all measures used to assess model performance and, if relevant, to compare multiple models. | 16 (22) | 11 (26) | 5 (15) | 5 (24) | 37 (22) |
| **e.** Describe any model updating (e.g., recalibration) arising from the validation, if done. | NA | 4/8 (50)* | 9/11 (82)* | 3/4 (75)* | 16/23 (70)* |

| | | | | | |
|---|---|---|---|---|---|
| **11. Risk groups**: Provide details on how risk groups were created, if done. | 20/22 (91)* | 13/15 (87)* | 18/20 (90)* | 12/13 (92)* | 63/70 (90)* |
| **12. Development vs. validation**: for validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors. | NA | 4 (9) | 0/17 (0)* | 5 (24) | 9/81 (11)* |
| ***Results*** | | | | | |
| **13. Participants**: | | | | | |
| **a.** Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful. | 29 (40) | 19 (44) | 14 (42) | 8 (38) | 70 (41) |
| **b.** Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome. | 18 (25) | 9 (21) | 4 (12) | 5 (24) | 36 (21) |
| **c.** For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome). | NA | 2 (5) | 19 (58) | 9 (43) | 30/97 (31)* |
| **14. Model development**: | | | | | |
| **a.** Specify the number of participants and outcome events in each analysis. | 47 (64) | NA | 22 (67) | 14 (67) | 83/127 (65)* |
| **b.** If done, report the unadjusted association between each candidate predictor and outcome. | 34/55 (62)* | NA | 14/25 (56)* | 11/14 (79)* | 59/94 (63)* |
| **15. Model specification**: | | | | | |
| **a.** Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point). | 15 (21) | NA | 1 (3) | 6 (29) | 22/127 (17)* |
| **b.** Explain how to the use the prediction model. | 26 (36) | NA | 5 (15) | 12 (57) | 43/127 (34)* |
| **16. Model performance**: report performance measures (with CIs) for the prediction model. | 7 (10) | 10 (23) | 3 (9) | 5 (24) | 25 (15) |

| | | | | | |
|---|---|---|---|---|---|
| **17. Model-updating**: if done, report the results from any model updating (i.e., model specification, model performance). | NA | 0/4 (0)* | NA | 1/3 (33)* | 1/7 (14)* |
| *Discussion* | | | | | |
| **18. Limitations**: discuss any limitations of the study (such as non-representative sample, few events per predictor, missing data). | 66 (90) | 36 (84) | 30 (91) | 18 (86) | 150 (88) |
| **19. Interpretation**: | | | | | |
| **a.** For validation, discuss the results with reference to performance in the development data, and any other validation data. | NA | 26 (61) | 19/29 (66)* | 13/20 (65)* | 58/92 (63)* |
| **b.** Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence. | 71 (97) | 40 (93) | 33 (100) | 20 (95) | 164 (97) |
| **20. Implications**: discuss the potential clinical use of the model and implications for future research. | 45 (62) | 21 (49) | 17 (52) | 17 (81) | 100 (59) |
| *Other information* | | | | | |
| **21. Supplementary information**: provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets. | 35 (48)† | 21 (49)† | 24 (73)† | 14 (67)† | 94 (55)† |
| **22. Funding**: give the source of funding and the role of the funders for the present study. | 17 (23) | 11 (26) | 9 (27) | 8 (38) | 45 (27) |

NA: not applicable (not all items of the TRIPOD statement are relevant to all types of prediction model studies)
Number of models for which an item was reported is shown with percentage in parentheses.
*Percentages are based on number of models for which that item was applicable (and should have been reported). Where this number deviates from the total number of models, the actual number of applicable models is presented as denominator.
†Item 21: number of models for which this item was applicable is unknown. It probably was applicable to all models that reported this item. Instead of presenting a percentage of 100, we based the percentage on the total number of models.



# Chapter 3

## Uniformity in measuring adherence to reporting guidelines: the example of TRIPOD for assessing completeness of reporting of prediction model studies

Pauline Heus
Johanna AAG Damen
Romin Pajouheshnia
Rob JPM Scholten
Johannes B Reitsma
Gary S Collins
Douglas G Altman
Karel GM Moons
Lotty Hooft

*BMJ Open 2019;9:e025611*

## Abstract

To promote uniformity in measuring adherence to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement, a reporting guideline for diagnostic and prognostic prediction model studies, and thereby facilitate comparability of future studies assessing its impact, we transformed the original 22 TRIPOD items into an adherence assessment form and defined adherence scoring rules.

TRIPOD specific challenges encountered were the existence of different types of prediction model studies and possible combinations of these within publications. More general issues included dealing with multiple reporting elements, reference to information in another publication, and nonapplicability of items.

We recommend our adherence assessment form to be used by anyone (e.g., researchers, reviewers, editors) evaluating adherence to TRIPOD, to make these assessments comparable. In general, when developing a form to assess adherence to a reporting guideline, we recommend formulating specific adherence elements (if needed multiple per reporting guideline item) using unambiguous wording and the consideration of issues of applicability in advance.

## Background

Incomplete reporting of research is considered to be a form of research waste.[1,2] To eventually implement research results in clinical guidelines and daily practice, one needs sufficient details regarding the research to critically appraise the methods and interpret study results in the context of existing evidence. [3-6]

To improve the reporting of health research, many reporting guidelines have been developed for various types of studies, such as the CONSORT (Consolidated Standards of Reporting Trials) statement, STARD (Standards for Reporting of Diagnostic Accuracy) statement, PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement, STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) statement , REMARK (REporting recommendations for tumour MARKer prognostic studies), and the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement.[7-15] A large number of reporting guidelines can be found on the website of the EQUATOR (Enhancing the QUAlity and Transparency Of health Research) Network, an international collaboration that supports the development and dissemination of reporting guidelines in order to achieve accurate, complete and transparent health research reporting (www.equator-network.org).[4,5]

Publishing a reporting guideline followed by some form of recommendation or journal endorsement is not enough for researchers to adhere to reporting guidelines - a more active implementation is usually required.[5] In their guidance for developers of health research reporting guidelines, Moher and colleagues proposed 18 steps to be taken in the development of a reporting guideline, including several post-publication activities.[6] One of these activities is to evaluate the actual adherence and thus use of a reporting guideline over time. Assessment of adherence has been carried out for CONSORT, STARD, and PRISMA.[16-23] In multiple evaluations of the same guideline different approaches to extract, score, and record adherence to items of the guideline were seen, making comparisons difficult.[17,21-23]. For example, a systematic review of studies assessing adherence to STARD found the number of items assessed was inconsistent and the criteria required for the reporting of an item to be complete differed between adherence evaluations. In addition, not all studies performed quantitative scoring, preventing an objective comparison of adherence between studies.[17,21-23] A systematic adherence-scoring-system is needed to enhance objectivity and to ensure consistent measurement of adherence to a reporting guideline. A unique assessment form for adherence evaluations would reduce variation in the number of items being evaluated, how multicomponent items are being handled, and the scoring rules (on item level and

overall adherence) applied, and thereby facilitate comparison of reporting between different fields and over time.

As the TRIPOD statement was only recently published (2015), its impact has not been assessed yet. However, recently a baseline measurement was performed to evaluate the extent to which prediction model studies before the introduction of TRIPOD reported each of the TRIPOD items.[24] Based on this, the TRIPOD steering committee aimed to develop a systematic and transparent adherence-scoring-system to be used by other researchers to facilitate and ensure uniformity in measuring adherence to TRIPOD in future studies. We also provide general recommendations on developing an adherence assessment form for other reporting guidelines.

## Development of the TRIPOD adherence assessment form

Our adherence assessment form contains all 22 main items of the original TRIPOD statement. Ten of these TRIPOD items actually comprise two (items 3, 4, 6, 7, 14, 15, and 19), three (items 5 and 13), or five (item 10) sub items (denoted by a, b, c, etc.; see Box 1).[15,25] For our TRIPOD adherence assessment form, we further specified these original TRIPOD items (main or sub items, hereafter referred to as items) into so-called adherence elements. When a TRIPOD item contains multiple elements to report, multiple adherence elements were used. For example, for TRIPOD item 5a *"Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres."* we defined three adherence elements to record information regarding 1) setting, 2) number, and 3) location of centres.

**Box 1. Items of the TRIPOD statement**

***Title and abstract***

1. **Title (D; V)**: identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.
2. **Abstract (D; V)**: provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.

***Introduction***

3. **Background and objectives**:
   - **a.** **(D; V)** Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.
   - **b.** **(D; V)** Specify the objectives, including whether the study describes the development or validation of the model or both.

*Methods*

4. **Source of data**:
   a. **(D; V)** Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.
   b. **(D; V)** Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.
5. **Participants**:
   a. **(D; V)** Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.
   b. **(D; V)** Describe eligibility criteria for participants.
   c. **(D; V)** Give details of treatments received, if relevant.
6. **Outcome**:
   a. **(D; V)** Clearly define the outcome that is predicted by the prediction model, including how and when assessed.
   b. **(D; V)** Report any actions to blind assessment of the outcome to be predicted.
7. **Predictors**:
   a. **(D; V)** Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.
   b. **(D; V)** Report any actions to blind assessment of predictors for the outcome and other predictors.
8. **Sample size (D; V)**: explain how the study size was arrived at.
9. **Missing data (D; V)**: Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.
10. **Statistical analysis methods**:
    a. **(D)** Describe how predictors were handled in the analyses.
    b. **(D)** Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.
    c. **(V)**For validation, describe how the predictions were calculated.
    d. **(D; V)** Specify all measures used to assess model performance and, if relevant, to compare multiple models.
    e. **(V)** Describe any model updating (e.g., recalibration) arising from the validation, if done.
11. **Risk groups (D; V)**: Provide details on how risk groups were created, if done.
12. **Development vs. validation (V)**: for validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.

*Results*

13. **Participants**:
    a. **(D; V)** Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.
    b. **(D; V)** Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.

**c.** **(V)** For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).

**14. Model development**:

**a.** **(D)** Specify the number of participants and outcome events in each analysis.

**b.** **(D)**If done, report the unadjusted association between each candidate predictor and outcome.

**15. Model specification**:

**a.** **(D)** Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).

**b.** **(D)** Explain how to the use the prediction model.

**16. Model performance (D;V)**: report performance measures (with CIs) for the prediction model.

**17. Model-updating (V)**: if done, report the results from any model updating (i.e., model specification, model performance).

***Discussion***

**18. Limitations (D;V)**: discuss any limitations of the study (such as non-representative sample, few events per predictor, missing data).

**19. Interpretation**:

**a.** **(V)** For validation, discuss the results with reference to performance in the development data, and any other validation data.

**b.** **(D;V)**Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.

**20. Implications (D;V)**: discuss the potential clinical use of the model and implications for future research.

***Other information***

**21. Supplementary information (D;V)**: provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.

**22. Funding (D;V)**: give the source of funding and the role of the funders for the present study.

*D;V*: item relevant to both development and external validation; *D*: item only relevant to development; *V*: item only relevant to external validation

We further distinguished four types of prediction model studies: model development, external validation, incremental value of adding one or more predictor(s) to an existing model, or a combination of development and external validation of the same model. Six TRIPOD items only apply to development of a prediction model (10a, 10b, 14a, 14b, 15a, and 15b) and six only to external validation (10c, 10e, 12, 13c, 17, and 19a) (Box 1).[15,25] All TRIPOD items, except for TRIPOD item 17, were considered applicable to incremental value reports. As not all TRIPOD items apply to all four types of prediction model studies, we defined four versions of the adherence assessment form, depending on whether a report described model development, external validation, a combination of these, or incremental value. If a report addresses both the development and external

validation of the same prediction model, the reporting of either should be assessed separately, and subsequently be combined for each adherence element.

There were several stages in the process of developing the adherence assessment form (Figure 1). All authors commented upon the first version of the form. A revised version was then piloted by four authors representing the TRIPOD steering committee (JBR, GSC, DGA, and KGMM). Based on their experiences adaptations to the form were made, mainly in the number and wording of the adherence assessment elements. Subsequently, the form was piloted by a group of various end-users consisting of PhD students, junior researchers, assistant and associate professors, professors and senior editors (n=16 ). Thereafter, three other authors (PH, JAAGD, RP) used the next version of the form when assessing six studies in duplicate. Items that led to disagreement or uncertainty more than once (items 2, 4b, 5a, 5c, 6a, 6b, 7b, 8, 10a, 10b, 10d, 11, 13a, 13b, 19 and 20) were discussed within the entire author team, leading to the final version of the form that was used to assess adherence to TRIPOD in a set of 146 publications.[24] The form was also used by another group assessing adherence to TRIPOD in prognostic models for diabetes (publication in preparation). Challenges encountered and discussions held in this stage, only led to textual refinements to the form. Our final adherence assessment form, including considerations and guidance regarding scoring and calculations, is summarised in Supplementary file 1. It can also be found on the website of the TRIPOD statement (www.tripod-statement.org/).

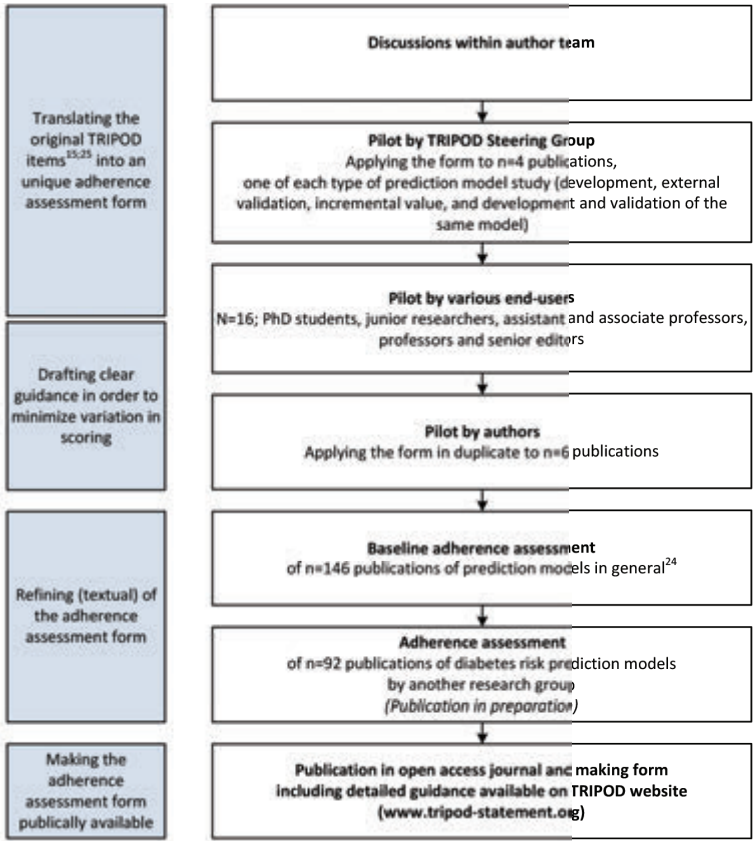


## Using the TRIPOD adherence assessment form

### Scoring adherence per TRIPOD item

First, one has to judge for each adherence element whether the requested information is available in a report. The elements are formulated as statements that can be answered with "yes " or "no" (see Supplementary file 1). For some elements it may be acceptable if authors in their report make explicit reference to another publication (i.e. explicitly mention that the information of that adherence element is described somewhere else). This is denoted by the answer option "referenced". For adherence elements that do not apply to a specific situation (for example reporting of follow-up (item 4b) might be not relevant in a diagnostic prediction model study), there is the answer option "not applicable".

The next step is to determine the adherence of a report per TRIPOD item. In general, if the answer to all adherence elements of a particular TRIPOD item is scored "yes" or "not applicable", the TRIPOD item is considered as adhered. In some situations a different scoring rule is used, which is described in the adherence assessment form for the corresponding items.




After each stage (except for the last one) the form was adapted and further refined.

**Figure 1. Process of developing the TRIPOD adherence assessment form with the aim of reducing unnecessary variation in scoring quality of reporting of prediction model studies based on TRIPOD**

### Overall adherence to TRIPOD

A report's overall TRIPOD adherence score is calculated by dividing the sum of the adhered TRIPOD items by the total number of applicable TRIPOD items. Since some TRIPOD items are not applicable to all four types of prediction model studies, this total varies. The total number of applicable TRIPOD items for development is 30, for external validation 30, for incremental value 35, and for development and development of the same model 36. In addition, five TRIPOD items (5c, 10e, 11, 14b, and 17) might not be applicable for specific reports (Supplementary file 1).

If one reviews multiple prediction model studies on their adherence to TRIPOD, overall adherence per TRIPOD item can be calculated by dividing the number of studies that

adhered to a specific TRIPOD item by the number of studies in which the specific TRIPOD item was applicable.

## Recommendations for developing and using a standardized form for assessing adherence to a reporting guideline

As described earlier, during the process of designing this adherence assessment form we extensively discussed, piloted, and refined our methods. One issue specific to TRIPOD we discussed, are the different types of prediction model studies (development, external validation, and incremental value) that can be found in various combinations within publications. As not all TRIPOD items apply to all types of prediction model studies, overall adherence scores need to be calculated per type of prediction model study.



A more general issue is how to deal with items containing several reporting elements. For TRIPOD we decided to determine adherence to a specific item by requiring complete information on all elements of that item. Hence, we created multiple adherence elements per TRIPOD item, as necessary.

Another issue with regard to scoring adherence is how to handle (elements of) TRIPOD items that were not applicable for a specific prediction model study. This not only concerns the judgements at the level of adherence elements, but also the calculations of adherence per TRIPOD item and of the overall adherence. Overall adherence, in the form of a percentage of items adhered to, requires a clear denominator of total number of items one can adhere to. One has to decide whether to take items that are considered not applicable into account in the numerator as well as in the denominator. Determining applicability is subjective and requires interpretation. In our experience, items for which interpretation was needed, sometimes indicated by phrases like 'if relevant' or 'if applicable', were the most difficult ones to score and these items are a potential threat to inter-assessor agreement.

We present our recommendations for developing and using a standardized form for measuring adherence to a reporting guideline in Box 2.

**Box 2. Recommendations for developing and using a standardized form for measuring adherence to a reporting guideline**

- Decide which items are applicable to the set of publications of which you are going to measure adherence to the reporting guideline.
- Split items of a reporting guideline that consist of several sub items and elements into separate adherence elements to enable more detailed judgment of reporting.
- Pay attention on explicit wording of adherence elements, to make them as objective as possible.
- Determine for which items reference to information in another publication (instead of explicit reporting of that information) is acceptable for adherence.
- Define how to handle items that are not applicable to a specific report:
  - agree on which items this may concern and in what specific situations a adherence element or item could be considered as not applicable;
  - decide how to incorporate the 'not applicable scores' in determining adherence, per item as well as overall.
- Provide the final tailored adherence assessment form with clear guidance about the procedure and pilot the document in a small number of studies with several assessors:
  - if there is poor agreement, discuss and refine the document;
  - with good agreement, complete the assessment for all publications.
- Abstract and document information separately for each adherence element. This creates flexibility, as one is able to decide post hoc which elements to incorporate in calculating adherence per item, and thus overall adherence.

## Concluding remarks

Evaluation of the impact of a reporting guideline should be as standardized and uniform as possible. However, this is not straightforward as reporting guidelines are usually not developed as an instrument to measure completeness of reporting. We presented an adherence assessment form that facilitates uniformity in measuring adherence to TRIPOD. The form is provided in Supplementary file 1 and on the website of the TRIPOD statement (www.tripod-statement.org). Although, when developing the form, we had researchers evaluating quality of reporting in mind as target users, it can also be used by others interested in assessing adherence to TRIPOD, like authors, journal reviewers, and editors. We would like to emphasize that our form should be used for assessing adherence to TRIPOD and not for assessing quality of prediction model studies (for which the Prediction model study Risk Of Bias Assessment Tool [PROBAST] is being developed).[26]

We did not perform formal user testing or reliability assessments, however we refined our adherence assessment form based on extensive discussions and pilot assessments within the author team, as well as by other potential users.

We advise developers of reporting guidelines to consider adherence issues and impact evaluation early in the process of guideline development, as also recommended by Moher and colleagues.[6] More specifically, attention should be paid to explicit wording of items, to make them as objective as possible and facilitate the interpretation of applicability and relevance.

## References


1. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374(9683):86-89.
2. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267-276.
3. Glasziou P, Meats E, Heneghan C, Shepperd S. What is missing from descriptions of treatment in trials and reviews? *BMJ*. 2008;336(7659):1472-1474.
4. Simera I, Altman DG, Moher D, Schulz KF, Hoey J. Guidelines for reporting health research: the EQUATOR network's survey of guideline authors. *PLoS Med*. 2008;5(6):e139.
5. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med*. 2010;8:24.
6. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7(2):e1000217.
7. Begg C, Cho M, Eastwood S, Horton R, Moher D, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA*. 1996;276(8):637-639.
8. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357(9263):1191-1194.
9. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152(11):726-732.
10. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
11. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med*. 2003;138(1):40-44.
12. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
13. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573-577.
14. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer*. 2005;93(4):387-391.

15. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55-63.
16. Smidt N, Rutjes AW, van der Windt DA, Ostelo RW, Bossuyt PM, et al. The quality of diagnostic accuracy studies since the STARD statement: has it improved? *Neurology.* 2006;67(5):792-797.
17. Korevaar DA, van Enst WA, Spijker R, Bossuyt PM, Hooft L. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evid Based Med.* 2014;19(2):47-54.
18. Korevaar DA, Wang J, van Enst WA, Leeflang MM, Hooft L, et al. Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. *Radiology.* 2015;274(3):781-789.
19. Kane RL, Wang J, Garrard J. Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *J Clin Epidemiol.* 2007;60(3):241-249.
20. Hopewell S, Dutton S, Yu LM, Chan AW, Altman DG. The quality of reports of randomised trials in 2000 and 2006: comparative study of articles indexed in PubMed. *BMJ.* 2010;340:c723.
21. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev.* 2012;11:Mr000030.
22. Samaan Z, Mbuagbaw L, Kosa D, Borg Debono V, Dillenburg R, et al. A systematic scoping review of adherence to reporting guidelines in health care literature. *J Multidiscip Healthc.* 2013;6:169-188.
23. Pussegoda K, Turner L, Garritty C, Mayhew A, Skidmore B, et al. Systematic review adherence to methodological or reporting quality. *Syst Rev.* 2017;6(1):131.
24. Heus P, Damen JAAG, Pajouheshnia R, Scholten RJPM, Reitsma JB, et al. Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. *BMC Med.* 2018;16(1):120.
25. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1-73.
26. Prediction model study Risk Of Bias Assessment Tool [PROBAST]. http://s371539711.initial-website.co.uk/probast/.

## Supplementary File
## Assessing adherence of prediction model reports to the TRIPOD guideline

This document provides guidance for extracting the relevant information and calculating summary scores to determine adherence of primary prediction model reports to the TRIPOD (Transparent Reporting of studies on prediction models for Individual Prognosis Or Diagnosis) reporting guideline (issued in January 2015; www.tripod-statement.org). **To be able to compare TRIPOD adherence evaluations, e.g. over time or over clinical domains, it is crucial that investigators use uniform methods, i.e. this adherence assessment form. If investigators decide to deviate from this form and scoring rules, they should be explicit and transparent about the changes they make.**

### Extracting the data

This TRIPOD adherence assessment form consists of two parts. Part A is to extract general information from a publication about the development and/or validation of a diagnostic or prognostic prediction model, or about the assessment of the incremental value of one or more predictors on top of an existing prediction model. Part B lists all 22 main items of the original TRIPOD reporting guideline, of which ten were divided in sub items (denoted by a, b, c, etc.). Below, presented in bold and further referred to as the TRIPOD items. To properly assess adherence of a study report to the TRIPOD reporting items, we further specified these TRIPOD items into multiple so-called adherence elements (denoted by i, ii, iii, …) simply because the original TRIPOD items often mentioned multiple elements to report. Accordingly, the form below provides a comprehensive tool to look for the information deemed necessary by the TRIPOD reporting guideline to judge the adherence of reports to this guideline.

There are four columns in which information can be entered: one for reports about the development of a prediction model [D], one for reports on external validation of a prediction model [V], one for reports on the incremental value of predictor(s) to an existing prediction model [IV], and one for reports on the development plus external validation of the same model [D+V]. If a report addresses both the development and validation of the same model, then both columns D and V should be used to assess the reporting of the development and external validation, and, subsequently, column D+V to combine the information of these two. If a report addresses the development of a model and external validation of a different model, one can use the columns D and V to assess the reporting however, information should not be combined using column D+V. For publications in which more than one (different) prediction model

is developed or validated, scoring could be based on the model of interest (or most clearly reported model).

The adherence elements are formulated as statements, for which there are four potential answer options: yes (Y), no (N), referenced (R), and not applicable (NA). For some elements it may be acceptable if authors in their report specifically reference to another publication (i.e. explicitly mention that the information of that data extraction element is described somewhere else). This is denoted by the answer option "R". For adherence elements that do not apply to a specific situation, there is the answer option "NA".

Some TRIPOD items do not apply to all four types of prediction model studes, e.g., TRIPOD item 10a *"Describe how predictors were handled in the analyses"*, is not applicable when reporting about external validation, whereas TRIPOD item 10c *"For validation, describe how the predictions were calculated"* does not apply to the reporting of model development. In such instances we state 'not applicable' and grey shaded these adherence elements.



**Calculating adherence to TRIPOD**

First, adherence of a report is calculated per TRIPOD item. If the answer to all adherence elements of a particular TRIPOD item is scored "yes", adherence to that TRIPOD item is scored as "1", and non adherence as "0". In some situations a different scoring rule is used, which is described in the adherence assessment form below for the corresponding items.

Subsequently, a report's overall TRIPOD adherence score can be calculated. This is calculated by dividing the sum of the adhered TRIPOD items by the total number of applicable TRIPOD items for that report. This total can vary since some TRIPOD items may be not applicable to all four types of prediction model studies. The total number of applicable TRIPOD items for D studies is 30, for V 30, for D+V 36 and for IV 35.[1] In addition, five TRIPOD items (5c, 10e, 11, 14b, and 17) might not be applicable for specific reports.

If one reviews multiple prediction model studies on their adherence to TRIPOD, overall adherence per TRIPOD item can be calculated by dividing the number of studies that adhered to a specific TRIPOD item by the number of studies in which the specific TRIPOD item was applicable.

1 TRIPOD item 21 is not taken into account in the overall score in any of the four types of studies.

| A. GENERAL INFORMATION | |
|---|---|
| Study ID | |
| First author | |
| Publication year | |
| Title | |
| Journal | |
| Diagnostic or prognostic prediction model? | □ Diagnostic<br>□ Prognostic |
| Type of prediction model study *(multiple options possible)* | □ Development<br>□ External validation<br>*If both development and external validation:*<br>□ same model/score<br>□ different models/scores<br><br>□ Incremental value |

| B. TRIPOD ITEMS | | | [D] Development | [V] External validation | [IV] Incremental value | [D+V] Development and external validation (of same model) |
|---|---|---|---|---|---|---|
| **Title and abstract** | | | | | | |
| *It is suggested to score items 1 and 2 (Title and Abstract) <u>after</u> scoring items 3 to 22, as only after reading the whole publication it can be judged whether the reporting in the title and abstract is complete.* | | | | | | |
| **Title** | **1** | **Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.** | **Score 1 if all elements are scored as "Y"** | **Score 1 if all elements are scored as "Y"** | **Score 1 if all elements are scored as "Y"** | **Score 1 if all elements are scored as "Y"** |
| | i | The words developing/development, validation/ validating, incremental/added value (or synonyms) are reported in the title | Y / N | Y / N | Y / N | =Y if D1i=Y AND V1i=Y |
| | ii | The words prediction, risk prediction, prediction model, risk models, prognostic models, prognostic indices, risk scores (or synonyms) are reported in the title | Y / N | Y / N | Y / N | =Y if D1ii=Y OR V1ii=Y |
| | iii | The target population is reported in the title | Y / N | Y / N | Y / N | =Y if D1iii=Y OR V1iii=Y |
| | iv | The outcome to be predicted is reported in the title | Y / N | Y / N | Y / N | =Y if D1iv=Y OR V1iv=Y |
| **Abstract** | **2** | **Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.** | **Score 1 if all elements are scored as "Y" or "NA"** | **Score 1 if all elements are scored as "Y" or "NA"** | **Score 1 if all elements are scored as "Y" or "NA"** | **Score 1 if all elements are scored as "Y" or "NA"** |
| | i | The objectives are reported in the abstract | Y / N | Y / N | Y / N | =Y if D2i=Y AND V2i=Y |
| | ii | Sources of data are reported in the abstract<br>*E.g. Prospective cohort, registry data, RCT data.* | Y / N | Y / N | Y / N | =Y if D2ii=Y AND V2ii=Y |
| | iii | The setting is reported in the abstract<br>*E.g. Primary care, secondary care, general population, adult care, or paediatric care. The setting should be reported for both the development and validation datasets, if applicable.* | Y / N | Y / N | Y / N | =Y if D2iii=Y AND V2iii=Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| | iv | A general definition of the study participants is reported in the abstract<br>*E.g. patients with suspicion of certain disease, patients with a specific disease, or general eligibility criteria.* | Y / N | Y / N | Y / N | =Y if D2iv=Y AND V2iv=Y |
| | v | The overall sample size is reported in the abstract | Y / N | Y / N | Y / N | =Y if D2v=Y AND V2v=Y |
| | vi | The number of events (or % outcome together with overall sample size) is reported in the abstract<br>*If a continuous outcome was studied, score Not applicable* | Y / N / NA | Y / N / NA | Y / N / NA | =Y if (D2vi=Y AND V2vi=(Y OR NA)) OR (D2vi = (Y OR NA) AND V2vi=Y) =NA if D2vi=NA AND V2vi=NA |
| | vii | Predictors included in the final model are reported in the abstract. For validation studies of well-known models, at least the name/acronym of the validated model is reported<br>*Broad descriptions are sufficient, e.g. 'all information from patient history and physical examination'.*<br>*Check in the main text whether all predictors of the final model are indeed reported in the abstract.* | Y / N | Y / N | Y / N | =Y if D2vii=Y OR V2vii=Y |
| | viii | The outcome is reported in the abstract | Y / N | Y / N | Y / N | =Y if D2viii=Y AND V2viii=Y |
| | ix | Statistical methods are described in the abstract<br>*For model development, at least the type of statistical model should be reported. For validation studies a quote like "model's discrimination and calibration was assessed" is considered adequate. If done, methods of updating should be reported.* | Y / N | Y / N | Y / N | =Y if D2ix=Y AND V2ix=Y |
| | x | Results for model discrimination are reported in the abstract<br>*This should be reported separately for development and validation if a study includes both development and validation..* | Y / N | Y / N | Y / N | =Y if D2x=Y AND V2x=Y |
| | xi | Results for model calibration are reported in the abstract<br>*This should be reported separately for development and validation if a study includes both development and validation.* | Y / N | Y / N | Y / N | =Y if D2xi=Y AND V2xi=Y |
| | xii | Conclusions are reported in the abstract<br>*In publications addressing both model development and validation, there is no need for separate conclusions for both; one conclusion is sufficient.* | Y / N | Y / N | Y / N | =Y if D2xii=Y OR V2xii=Y |
| **Background and objectives** | **3a** | **Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** |
| | i | The background and rationale are presented | Y / N | Y / N | Y / N | =Y if D3ai=Y OR V3ai=Y |
| | ii | Reference to existing models is included (or stated that there are no existing models) | Y / N | Y / N | Y / N | =Y if D3aii=Y OR V3aii=Y |
| | **3b** | **Specify the objectives, including whether the study describes the development or validation of the model or both.** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** |
| | i | It is stated whether the study describes development and/or validation and/or incremental (added) value | Y / N | Y / N | Y / N | =Y if D3bi=Y AND V3bi=Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Methods** | | | | | | |
| **Source of data** | **4a** | **Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** |
| | i | The study design/source of data is described<br>*E.g. Prospectively designed, existing cohort, existing RCT, registry/medical records, case control, case series.*<br>*This needs to be explicitly reported; reference to this information in another article alone is insufficient.* | Y / N | Y / N | Y / N | =Y if D4ai=Y AND V4ai=Y |
| | **4b** | **Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.** | **Score 1 if all elements are scored as "Y", "NA", or "R"** | **Score 1 if all elements are scored as "Y", "NA", or "R"** | **Score 1 if all elements are scored as "Y", "NA", or "R"** | **Score 1 if all elements are scored as "Y", "NA", or "R"** |
| | i | The starting date of accrual is reported | Y / N / R | Y / N / R | Y / N / R | =Y if (D4bi=Y AND V4bi=(Y OR R)) OR (D4bi =(Y OR R) AND V4bi=Y) =R if D4bi=R AND V4bi=R |
| | ii | The end date of accrual is reported | Y / N / R | Y / N / R | Y / N / R | =Y if (D4bii=Y AND V4bii=(Y OR R)) OR (D4bii =(Y OR R) AND V4bii=Y) =R if D4bii=R AND V4bii=R |
| | iii | The length of follow-up <u>and</u> prediction horizon/time frame are reported, if applicable<br>*E.g. "Patients were followed from baseline for 10 years" and "10-year prediction of…"; notably for prognostic studies with long term follow-up.*<br>*If this is not applicable for an article (i.e. diagnostic study or no follow-up), then score Not applicable.* | Y / N / NA | Y / N / NA | Y / N / NA | =Y if (D4biii=Y AND V4biii=(Y OR NA)) OR (D4biii =(Y OR NA) AND V4biii=Y) =NA if D4biii=NA AND V4biii=NA |
| **Participants** | **5a** | **Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.** | **Score 1 if all elements are scored as "Y" or "R"** | **Score 1 if all elements are scored as "Y" or "R"** | **Score 1 if all elements are scored as "Y" or "R"** | **Score 1 if all elements are scored as "Y" or "R"** |
| | i | The study setting is reported (e.g. primary care, secondary care, general population)<br>*E.g.: 'surgery for endometrial cancer patients' is considered to be enough information about the study setting.* | Y / N / R | Y / N / R | Y / N / R | =Y if (D5ai=Y AND V5ai=(Y OR R)) OR (D5ai =(Y OR R) AND V5ai=Y) =R if D5ai=R AND V5ai=R |
| | ii | The number of centres involved is reported<br>*If the number is not reported explicitly, but can be concluded from the name of the centre/centres, or if clearly a single centre study, score Yes.* | Y / N / R | Y / N / R | Y / N / R | =Y if (D5aii=Y AND V5aii=(Y OR R)) OR (D5aii =(Y OR R) AND V5aii=Y) =R if D5aii=R AND V5aii=R |

| | | | | | | |
|---|---|---|---|---|---|---|
| | iii | The geographical location (at least country) of centres involved is reported<br>*If no geographical location is specified, but the location can be concluded from the name of the centre(s), score Yes.* | Y / N / R | Y / N / R | Y / N / R | =Y if (D5aiii=Y AND V5aiii=(Y OR R)) OR (D5aiii =(Y OR R) AND V5aiii=Y) =R if D5aiii=R AND V5aiii=R |
| | **5b** | **Describe eligibility criteria for participants.** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** |
| | i | In-/exclusion criteria are stated<br>*These should explicitly be stated. Reasons for exclusion only described in a patient flow is not sufficient.* | Y / N | Y / N | Y / N | =Y if D5bi=Y AND V5bi=Y |
| | **5c** | **Give details of treatments received, if relevant.** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** |
| | i | Details of any treatments received are described<br>*This item is notably for prognostic modelling studies and is about treatment at baseline or during follow-up. The 'if relevant' judgment of treatment requires clinical knowledge and interpretation.*<br>*If you are certain that treatment was not relevant, e.g. in some diagnostic model studies, score Not applicable* | Y / N / NA | Y / N / NA | Y / N / NA | =Y if (D5ci=Y AND V5ci=(Y OR NA) OR (D5ci=(Y OR NA) AND V5ci=Y) =NA if D5ci=NA AND V5ci=NA |
| **Outcome** | **6a** | **Clearly define the outcome that is predicted by the prediction model, including how and when assessed.** | **Score 1 if all elements are scored as "Y" or "R"** | **Score 1 if all elements are scored as "Y" or "R"** | **Score 1 if all elements are scored as "Y" or "R"** | **Score 1 if all elements are scored as "Y" or "R"** |
| | i | The outcome definition is clearly presented<br>*This should be reported separately for development and validation if a publication includes both.* | Y / N / R | Y / N / R | Y / N / R | =Y if (D6ai=Y AND V6ai=(Y OR R)) OR (D6ai =(Y OR R) AND V6ai=Y) =R if D6ai=R AND V6ai=R |
| | ii | It is described how outcome was assessed (including all elements of any composite, for example CVD [e.g. MI, HF, stroke]). | Y / N / R | Y / N / R | Y / N / R | =Y if (D6aii=Y AND V6aii=(Y OR R)) OR (D6aii =(Y OR R) AND V6aii=Y) =R if D6aii=R AND V6aii=R |
| | iii | It is described when the outcome was assessed (time point(s) since T0) | Y / N / R | Y / N / R | Y / N / R | =Y if (D6aiii=Y AND V6aiii=(Y OR R)) OR (D6aiii =(Y OR R) AND V6aiii=Y) =R if D6aiii=R AND V6aiii=R |
| | **6b** | **Report any actions to blind assessment of the outcome to be predicted.** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | i | Actions to blind assessment of outcome to be predicted are reported<br>*If it is clearly a non-issue (e.g. all-cause mortality or an outcome not requiring interpretation), score Yes. In all other instances, an explicit mention is expected.* | Y / N | Y / N | Y / N | =Y if D6bi=Y AND V6bi=Y |
| **Predictors** | **7a** | **Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.** | **Score 1 if all elements are scored as "Y" or "R"** | **Score 1 if all elements are scored as "Y" or "R"** | **Score 1 if all elements are scored as "Y" or "R"** | **Score 1 if all elements are scored as "Y" or "R"** |
| | i | All predictors are reported<br>*For development, "all predictors" refers to all predictors that potentially could have been included in the 'final' model (including those considered in any univariable analyses). For validation, "all predictors" means the predictors in the model being evaluated.* | Y / N | Y / N | Y / N | =Y if D7ai=Y |
| | ii | Predictor definitions are clearly presented | Y / N / R | Y / N / R | Y / N / R | =Y if (D7aii=Y AND V7aii=(Y OR R)) OR (D7aii =(Y OR R) AND V7aii=Y) =R if D7aii=R AND V7aii=R |
| | iii | It is clearly described how the predictors were measured | Y / N / R | Y / N / R | Y / N / R | =Y if (D7aiii=Y AND V7aiii=(Y OR R)) OR (D7aiii =(Y OR R) AND V7aiii=Y) =R if D7aiii=R AND V7aiii=R |
| | iv | It is clearly described when the predictors were measured | Y / N / R | Y / N / R | Y / N / R | =Y if (D7aiv=Y AND V7aiv=(Y OR R)) OR (D7aiv=(Y OR R) AND V7aiv=Y) =R if D7aiv=R AND V7aiv=R |
| | **7b** | **Report any actions to blind assessment of predictors for the outcome and other predictors.** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** |
| | i | It is clearly described whether predictor assessments were blinded for outcome<br>*For predictors for which it is clearly a non-issue (e.g. automatic blood pressure measurement, age, sex) and for instances where the predictors were clearly assessed before outcome assessment, score Yes. For all other predictors an explicit mention is expected.* | Y / N | Y / N | Y / N | =Y if D7bi=Y AND V7bi=Y |
| | ii | It is clearly described whether predictor assessments were blinded for the other predictors | Y / N | Y / N | Y / N | =Y if D7bii=Y AND V7bii=Y |
| **Sample size** | **8** | **Explain how the study size was arrived at.** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** |
| | i | It is explained how the study size was arrived at<br>*Is there any mention of sample size, e.g. whether this was done on statistical grounds or practical/logistical grounds (e.g. an existing study cohort or data set of a RCT was used)?* | Y / N | Y / N | Y / N | =Y if D8i=Y AND V8i=Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Missing data** | **9** | **Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.** | **Score 1 if all elements are scored as "Y" or "NA"** | **Score 1 if all elements are scored as "Y" or "NA"** | **Score 1 if all elements are scored as "Y" or "NA"** | **Score 1 if all elements are scored as "Y" or "NA"** |
| | i | The method for handling missing data (predictors and outcome) is mentioned<br>*E.g. Complete case (explicit mention that individuals with missing values have been excluded), single imputation, multiple imputation, mean/median imputation.*<br>*If there is no missing data, there should be an explicit mention that there is no missing data for all predictors and outcome. If so, score Yes.*<br>*If it is unclear whether there is missing data (from e.g. the reported methods or results), score No.*<br>*If it is clear there is missing data, but the method for handling missing data is unclear, score No.* | Y / N | Y / N | Y / N | =Y if D9i=Y AND V9i=Y |
| | ii | If missing data were imputed, details of the software used are given<br>*When under 9i explicit mentioning of no missing data, complete case analysis or no imputation applied, score Not applicable* | Y / N / NA | Y / N / NA | Y / N / NA | =Y if (D9ii=Y AND V9ii=(Y OR NA)) OR (D9ii =(Y OR NA) AND V9ii=Y) =NA if D9ii=NA AND V9ii=NA |
| | iii | If missing data were imputed, a description of which variables were included in the imputation procedure is given.<br>*When under 9i explicit mentioning of no missing data, complete case analysis or no imputation applied, score Not applicable* | Y / N / NA | Y / N / NA | Y / N / NA | =Y if (D9iii=Y AND V9iii=(Y OR NA)) OR (D9iii =(Y OR NA) AND V9iii=Y) =NA if D9iii=NA AND V9iii=NA |
| | iv | If multiple imputation was used, the number of imputations is reported<br>*When under 9i explicit mentioning of no missing data, complete case analysis or no imputation applied, score Not applicable* | Y / N / NA | Y / N / NA | Y / N / NA | =Y if (D9iv=Y AND V9iv=(Y OR NA)) OR (D9iv=(Y OR NA) AND V9iv=Y) =NA if D9iv=NA AND V9iv=NA |
| **Statistical analysis methods** | **10a** | **Describe how predictors were handled in the analyses.** | **Score 1 if all elements are scored as "Y" or "NA"** | **Not applicable** | **Score 1 if all elements are scored as "Y" or "NA"** | **Score 1 if all elements are scored as "Y" or "NA"** |
| | i | For continuous predictors it is described whether they were modelled as linear, nonlinear (type of transformation specified) or categorized<br>*A general statement is sufficient, no need to describe this for each predictor separately.*<br>*If no continuous predictors were reported, score Not applicable.* | Y / N / NA | Not applicable | Y / N / NA | =D10ai |
| | ii | For categorical or categorized predictors, the cut-points were reported<br>*If no categorical or categorized predictors were reported, score Not applicable.* | Y / N / NA | Not applicable | Y / N / NA | = D10aii |
| | iii | For categorized predictors the method to choose the cut-points was clearly described<br>*If no categorized predictors, score Not applicable.* | Y / N / NA | Not applicable | Y / N / NA | = D10aiii |

| | | | | | |
|---|---|---|---|---|---|
| **10b** | **Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.** | **Score 1 if all elements are scored as "Y" or "NA"** | **Not applicable** | **Score 1 if all elements are scored as "Y" or "NA"** | **Score 1 if all elements are scored as "Y" or "NA"** |
| i | The type of statistical model is reported<br>*E.g. Logistic, Cox, other regression model (e.g. Weibull, ordinal), other statistical modelling (e.g. neural network)* | Y / N | Not applicable | Y / N | =D10bi |
| ii | The approach used for predictor selection <u>before</u> modelling is described<br>*'Before modelling' means before any univariable or multivariable analysis of predictor-outcome associations.*<br>*If no predictor selection before modelling is done, score Not applicable.*<br>*If it is unclear whether predictor selection before modelling is done, score No.*<br>*If it is clear there was predictor selection before modelling but the method was not described, score No.* | Y / N / NA | Not applicable | Y / N / NA | = D10bii |
| iii | The approach used for predictor selection <u>during</u> modelling is described<br>*E.g. Univariable analysis, stepwise selection, bootstrap, Lasso.*<br>*'During modelling' includes both univariable or multivariable analysis of predictor-outcome associations.*<br>*If no predictor selection during modelling is done (so-called full model approach), score Not applicable.*<br>*If it is unclear whether predictor selection during modelling is done, score No.*<br>*If it is clear there was predictor selection during modelling but the method was not described, score No.* | Y / N / NA | Not applicable | Y / N / NA | = D10biii |
| iv | Testing of interaction terms is described<br>*If it is explicitly mentioned that interaction terms were not addressed in the prediction model, score Yes.*<br>*If interaction terms were included in the prediction model, but the testing is not described, score No.* | Y / N | | Y / N | =D10biv |
| v | Testing of the proportionality of hazards in survival models is described<br>*If no proportional hazard model is used, score Not applicable.* | Y / N / NA | Not applicable | Y / N / NA | =D10bv |
| vi | Internal validation is reported<br>*E.g. Bootstrapping, cross validation, split sample.*<br>*If the use of internal validation is clearly a non-issue (e.g. in case of very large data sets), score Yes. For all other situations an explicit mention is expected.* | Y / N | Not applicable | Y / N | =D10bvi |
| **10c** | **For validation, describe how the predictions were calculated.** | **Not applicable** | **Score 1 if extraction item is scored as "Y"** | **Score 1 if extraction item is scored as "Y"** | **Score 1 if extraction item is scored as "Y"** |
| i. | It is described how predictions for individuals (in the validation set) were obtained from the model being validated<br>*E.g. Using the original reported model coefficients with or without the intercept, and/or using updated or refitted model coefficients, or using a nomogram, spreadsheet or web calculator.* | Not applicable | Y / N | Y / N | =V10ci |

| | | | | | | |
|---|---|---|---|---|---|---|
| | **10d** | **Specify all measures used to assess model performance and, if relevant, to compare multiple models.[2]**<br>*These should be described in the methods section of the paper (item 16 addresses the reporting of the results for model performance).* | **Score 1 if elements 10di and 10dii are scored as "Y"[2]** | **Score 1 if elements 10di and 10dii are scored as "Y"[2]** | **Score 1 if all elements are scored as "Y"[2]** | **Score 1 if elements 10di and 10dii are scored as "Y"[2]** |
| | i | Measures for model discrimination are described<br>*E.g. C-index / area under the ROC curve* | Y / N | Y / N | Y / N | =Y if D10di=Y AND V10di=Y |
| | ii | Measures for model calibration are described<br>*E.g. calibration plot, calibration slope or intercept, calibration table, Hosmer Lemeshow test, O/E ratio.* | Y / N | Y / N | Y / N | =Y if D10dii=Y AND V10dii=Y |
| | iii | Other performance measures are described<br>*E.g. $R^2$, Brier score, predictive values, sensitivity, specificity, AUC difference, decision curve analysis, net reclassification improvement, integrated discrimination improvement, AIC* | Y / N | Y / N | Y / N | =Y if D10diii=Y AND V10diii=Y |
| | **10e** | **Describe any model updating (e.g., recalibration) arising from the validation, if done.** | **Not applicable** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** |
| | i | A description of model-updating is given<br>*E.g. Intercept recalibration, regression coefficient recalibration, refitting the whole model, adding a new predictor*<br>*If updating was done, it should be clear which updating method was applied to score Yes.*<br>*If it is not explicitly mentioned that updating was applied in the study, score this item as 'Not applicable'.* | Not applicable | Y / N / NA | Y / N / NA | =V10ei |
| **Risk groups** | **11** | **Provide details on how risk groups were created, if done.** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** |
| | i | If risk groups were created, risk group boundaries (risk thresholds) are specified<br>*Score this item separately for development and validation if a study includes both development and validation.*<br>*If risk groups were not created, score this item as not applicable.* | Y / N / NA | Y / N / NA | Y / N / NA | =Y if (D11i=Y AND V11i=(Y OR NA)) OR (D11i=(Y OR NA) AND V11i=Y) =NA if D11i=NA AND V11i=NA |

2 Discrimination and calibration are the two key aspects that characterize the performance of a prediction model and the TRIPOD guideline states that these two measures should be mentioned in every prediction model report. Various other measures of model performance can sometimes be reported (see examples provided at data extraction element 10diii). For reports on D and V and DV, we considered that discrimination and calibration had to be reported to adhere to item 10d. Other overall performance measures such as ($R^2$, Brier score or AIC) were not deemed essential for the scoring of overall adherence in D, V and D+V reports. For reports on the incremental value (IV reports) the reporting of other performance measures, like AUC difference or net reclassification improvement, were considered essential in addition to discrimination and calibration.

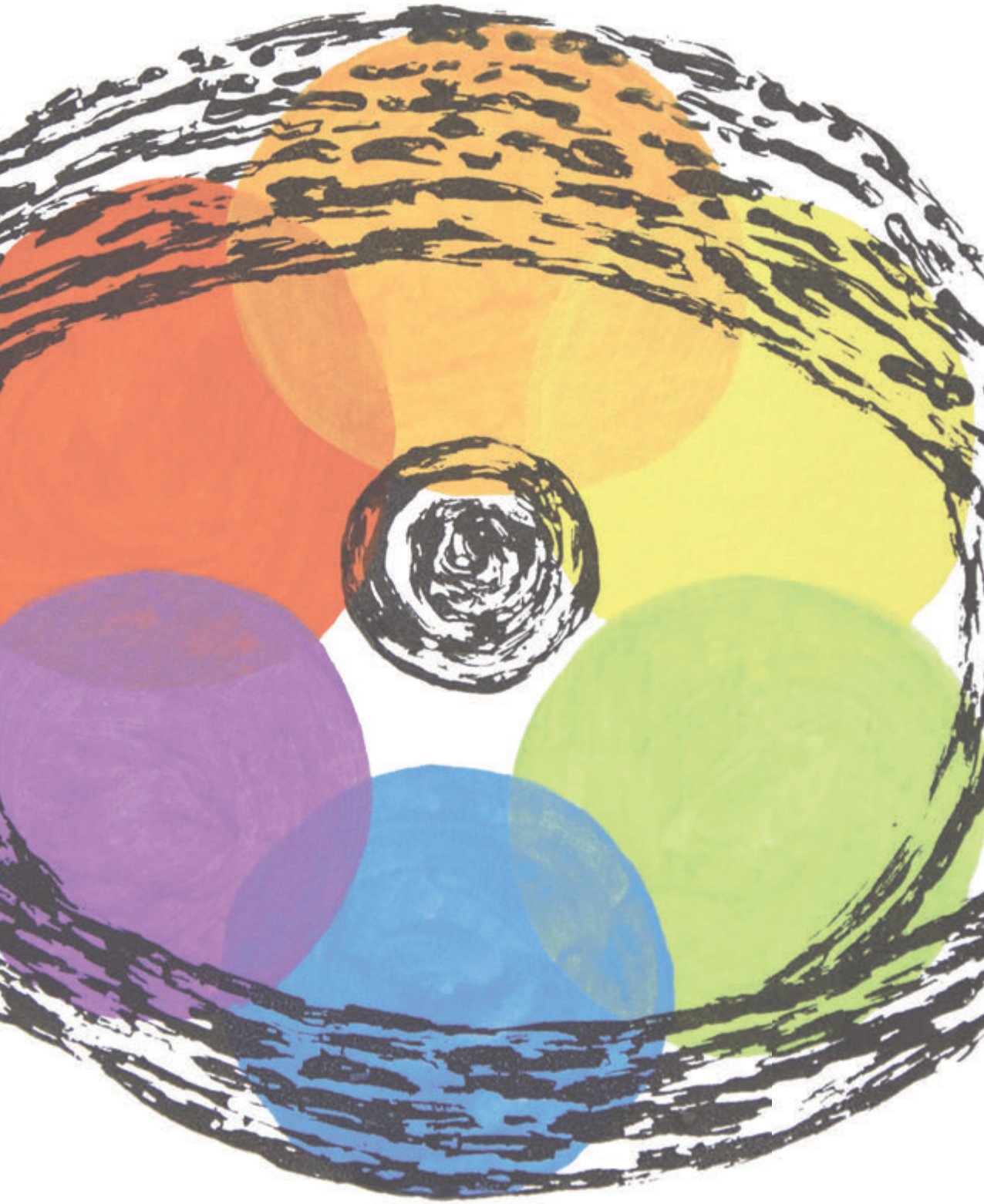
| | | | | | | |
|---|---|---|---|---|---|---|
| **Development vs. validation** | **12** | **For validation, identify any differences from the development data in setting, eligibility criteria, outcome and predictors.** | **Not applicable** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y" or "NA"** | **Score 1 if element is scored as "Y"** |
| | i | Differences or similarities in definitions with the development study are described<br>*Mentioning of any differences in all four (setting, eligibility criteria, predictors and outcome) is required to score Yes.*<br>*If it is explicitly mentioned that there were no differences in setting, eligibility criteria, predictors and outcomes, score Yes.*<br>*For incremental value reports, in case additional predictors are not added to a previously developed prediction model but rather added to conventional predictors in a newly fitted model, score Not applicable.* | Not applicable | Y / N | Y / N / NA | =V12i |
| **Results** | | | | | | |
| **Participants** | **13a** | **Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.** | **Score 1 if all elements are scored as "Y" or "NA"** | **Score 1 if the elements are scored as "Y" or "NA"** | **Score 1 if all elements are scored as "Y" or "NA"** | **Score 1 if all elements are scored as "Y" or "NA"** |
| | i | The flow of participants is reported | Y / N | Y / N | Y / N | =Y if D13ai=Y AND V13ai=Y |
| | ii | The number of participants with and without the outcome are reported<br>*If outcomes are continuous, score Not applicable.* | Y / N / NA | Y / N / NA | Y / N / NA | =Y if (D13aii=Y AND V13aii=(Y OR NA)) OR (D13aii =(Y OR NA) AND V13aii=Y) =NA if D13aii=NA AND V13aii=NA |
| | iii | A summary of follow-up time is presented<br>*This notably applies to prognosis studies and diagnostic studies with follow-up as diagnostic outcome.*<br>*If this is not applicable for an article (i.e. diagnostic study or no follow-up), then score Not applicable.* | Y / N / NA | Y / N / NA | Y / N / NA | =Y if (D13aiii=Y AND V13aiii=(Y OR NA)) OR (D13aiii =(Y OR NA) AND V13aiii=Y) =NA if D13aiii=NA AND V13aiii=NA |
| | **13b** | **Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.** | **Score 1 if all elements are scored as "Y"** | **Score 1 if all elements are scored as "Y"** | **Score 1 if all elements are scored as "Y"** | **Score 1 if all elements are scored as "Y"** |
| | i | Basic demographics are reported | Y / N | Y / N | Y / N | =Y if D13bi=Y AND V13bi=Y |
| | ii | Summary information is provided for all predictors included in the final developed/validated model | Y / N | Y / N | Y / N | =Y if D13bii=Y AND V13bii=Y |
| | iii | The number of participants with missing data for predictors is reported | Y / N | Y / N | Y / N | =Y if D13biii=Y AND V13biii=Y |
| | iv | The number of participants with missing data for the outcome is reported | Y / N | Y / N | Y / N | =Y if D13biv=Y AND V13biv=Y |
| | **13c** | **For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).** | **Not applicable** | **Score 1 if all elements are scored as "Y"** | **Score 1 if all elements are scored as "Y" or "NA"** | **Score 1 if all elements are scored as "Y"** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | i | Demographic characteristics (at least age and gender) of the validation study participants are reported along with those of the original development study<br>*For incremental value reports, in case additional predictors are not added to a previously developed prediction model but rather added to conventional predictors in a newly fitted model, score Not applicable.* | Not applicable | Y / N | Y / N / NA | =V13ci |
| | ii | Distributions of predictors in the model of the validation study participants are reported along with those of the original development study<br>*For incremental value reports, in case additional predictors are not added to a previously developed prediction model but rather added to conventional predictors in a newly fitted model, score Not applicable.* | Not applicable | Y / N | Y / N / NA | =V13cii |
| | iii | Outcomes of the validation study participants are reported along with those of the original development study<br>*For incremental value reports, in case additional predictors are not added to a previously developed prediction model but rather added to conventional predictors in a newly fitted model, score Not applicable.* | Not applicable | Y / N | Y / N / NA | =V13ciii |
| **Model development** | **14a** | **Specify the number of participants and outcome events in each analysis.** | **Score 1 if both elements are scored as "Y" or "NA"** | **Not applicable** | **Score 1 if both elements are scored as "Y" or "NA"** | **Score 1 if both elements are scored as "Y" or "NA"** |
| | i | The number of participants in each analysis (e.g. in the analysis of each model if more than one model is developed) is specified | Y / N | Not applicable | Y / N | =D14ai |
| | ii | The number of outcome events in each analysis is specified (e.g. in the analysis of each model if more than one model is developed)<br>*If outcomes are continuous, score Not applicable.* | Y / N / NA | Not applicable | Y / N / NA | =D14aii |
| | **14b** | **If done, report the unadjusted association between each candidate predictor and outcome.** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** | **Not applicable** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** | **Score 1 if element is scored as "Y"; score *Not applicable* if element is scored as "NA"** |
| | i | The unadjusted associations between each predictor and outcome are reported<br>*If any univariable analysis is mentioned in the methods but not in the results, score No.*<br>*If nothing on univariable analysis (in methods or results) is reported, score this item as Not applicable* | Y / N / NA | Not applicable | Y / N / NA | =D14bi |
| **Model specification** | **15a** | **Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).** | **Score 1 if both elements are scored as "Y"** | **Not applicable** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** |
| | i | The regression coefficient (or a derivative such as hazard ratio, odds ratio, risk ratio) for each predictor in the model is reported | Y / N | Not applicable | Y / N | =D15ai |
| | ii | The intercept or the cumulative baseline hazard (or baseline survival) for at least one time point is reported | Y / N | Not applicable | Y / N | =D15aii |

| | | | | | | |
|---|---|---|---|---|---|---|
| | **15b** | **Explain how to use the prediction model.** | **Score 1 if element is scored as "Y"** | **Not applicable** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** |
| | i | An explanation (e.g. a simplified scoring rule, chart, nomogram of the model, reference to online calculator, or worked example) is provided to explain how to use the model for individualised predictions. | Y / N | Not applicable | Y / N | =D15bi |
| **Model performance** | **16** | **Report performance measures (with confidence intervals) for the prediction model.[3]** *These should be described in results section of the paper (item 10 addresses the reporting of the methods for model performance).* | **Score 1 if elements 16i-16iii are scored as "Y"[3]** | **Score 1 if elements 16i-16iii are scored as "Y"[3]** | **Score 1 if all elements are scored as "Y"[3]** | **Score 1 if elements 16i-16iii are scored as "Y"** [3] |
| | i | A discrimination measure is presented *E.g. C-index / area under the ROC curve* | Y / N | Y / N | Y / N | =Y if D16i=Y AND V16i=Y |
| | ii | The confidence interval (or standard error) of the discrimination measure is presented | Y / N | Y / N | Y / N | =Y if D16ii=Y AND V16ii=Y |
| | iii | Measures for model calibration are described *E.g. calibration plot, calibration slope or intercept, calibration table, Hosmer Lemeshow test, O/E ratio.* | Y / N | Y / N | Y / N | =Y if D16iii=Y AND V16iii=Y |
| | iv | Other model performance measures are presented *E.g. $R^2$, Brier score, predictive values, sensitivity, specificity, AUC difference, decision curve analysis, net reclassification improvement, integrated discrimination improvement, AIC.* | Y / N | Y / N | Y / N | =Y if D16iv=Y AND V16iv=Y |
| **Model updating** | **17** | **If done, report the results from any model updating (i.e., model specification, model performance, recalibration).** *If updating was not done, score this TRIPOD item as 'Not applicable'.* | **Not applicable** | **Score 1 if all elements are scored as "Y"** | **Not applicable** | **Score 1 if all elements are scored as "Y"** |
| | i | The updated regression coefficients for each predictor in the model are reported *If model updating was described as 'not needed', score Yes.* | Not applicable | Y / N | Not applicable | =V17i |
| | ii | The updated intercept or cumulative baseline hazard or baseline survival (for at least one time point) is reported *If model updating was described as 'not needed', score Yes.* | Not applicable | Y / N | Not applicable | =V17ii |
| | iii | The discrimination of the updated model is reported | Not applicable | Y / N | Not applicable | =V17iii |
| | iv | The confidence interval (or standard error) of the discrimination measure of the updated model is reported | Not applicable | Y / N | Not applicable | =V17iv |
| | v | The calibration of the updated model is reported | Not applicable | Y / N | Not applicable | =V17v |
| **Discussion** | | | | | | |
| **Limitations** | **18** | **Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** |
| | i | Limitations of the study are discussed *Stating any limitation is sufficient.* | Y / N | Y / N | Y / N | =Y if D18i=Y OR V18i=Y |

3 See also footnote 2. Discrimination and calibration are the two key aspects that characterize the performance of a prediction model and the TRIPOD guideline states that these two measures should be reported in every prediction model report. Various other measures of model performance can sometimes be reported (see examples provided at data extraction element 16iv). For reports on D and V and D+V, we considered that discrimination and calibration had to be reported to adhere to item 16. Other overall performance measures such as ($R^2$, Brier score or AIC) were not deemed essential for the scoring of overall adherence in D, V and D+V reports. For reports on the incremental value (IV reports) the reporting of other performance measures, like AUC difference or net reclassification improvement, were considered essential in addition to discrimination and calibration.

| | | | | | | |
|---|---|---|---|---|---|---|
| **Interpretation** | **19a** | **For validation, discuss the results with reference to performance in the development data, and any other validation data.** | **Not applicable** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** |
| | i | Comparison of results to reported performance in development studies and/or other validation studies is given | Not applicable | Y / N | Y / N | =V19ai |
| | **19b** | **Give an overall interpretation of the results considering objectives, limitations, results from similar studies and other relevant evidence.** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** | **Score 1 if element is scored as "Y"** |
| | i | An overall interpretation of the results is given | Y / N | Y / N | Y / N | =Y if D19bi=Y OR V19bi=Y |
| **Implications** | **20** | **Discuss the potential clinical use of the model and implications for future research.** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** |
| | i | The potential clinical use is discussed<br>*E.g. an explicit description of the context in which the prediction model is to be used (e.g. to identify high risk groups to help direct treatment, or to triage patients for referral to subsequent care).* | Y / N | Y / N | Y / N | =Y if D20i=Y OR V20i=Y |
| | ii | Implications for future research are discussed<br>*E.g. a description of what the next stage of investigation of the prediction model should be, such as "We suggest further external validation".* | Y / N | Y / N | Y / N | =Y if D20ii=Y OR V20ii=Y |
| **Other information** | | | | | | |
| **Supplementary information** | **21** | **Provide information about the availability of supplementary resources, such as study protocol, web calculator, and data sets.** | **Not included in overall scoring** | **Not included in overall scoring** | **Not included in overall scoring** | **Not included in overall scoring** |
| | i | Information about supplementary resources is provided | Y / N | Y / N | Y / N | =Y if D21i=Y OR V21i=Y |
| **Funding** | **22** | **Give the source of funding and the role of the funders for the present study.** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** | **Score 1 if both elements are scored as "Y"** |
| | i | The source of funding is reported or there is explicit mention that there was no external funding involved | Y / N | Y / N | Y / N | =Y if D22i=Y OR V22i=Y |
| | ii | The role of funders is reported or there is explicit mention that there was no external funding | Y / N | Y / N | Y / N | =Y if D22ii=Y OR V22ii=Y |



# Chapter 4

## Transparent reporting of multivariable prediction models in journal and conference abstracts: TRIPOD for Abstracts

Pauline Heus
Johannes B Reitsma
Gary S Collins
Johanna AAG Damen
Rob JPM Scholten
Douglas G Altman
Karel GM Moons
Lotty Hooft

*Provisionally accepted for publication in the Annals of Internal Medicine*

## Abstract

Clear and informative reporting in titles and abstracts is essential to help readers and reviewers identify potentially relevant studies and give them the information they need to decide whether to read the full text. Although the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement provides general recommendations for reporting titles and abstracts, more detailed guidance appears to be desirable. We present TRIPOD for Abstracts, a checklist and corresponding guidance for reporting diagnostic or prognostic prediction model studies in journal and conference abstracts.

We first established a list of 32 potentially relevant items for inclusion in TRIPOD for Abstracts from TRIPOD and other reporting guidelines for abstracts. This list served as the basis for a modified Delphi procedure conducted as a web-based survey. Of 110 experts in prediction modeling invited to take part in the survey, 71 (65%) participated.

After two Delphi rounds, 21 items were agreed as essential when reporting prediction model studies in abstracts. In the third round, the participants were asked to provide feedback on a draft version of TRIPOD for Abstracts. Following their suggestions, items were combined and an item on protocol availability was added.

The final TRIPOD for Abstracts checklist contains 12 items and is applicable to journal and conference abstracts that describe the development or external validation of a diagnostic or prognostic prediction model, or describing the added value of predictors to an existing prediction model, regardless the clinical domain or statistical approach used (including artificial intelligence and machine learning approaches).

## Introduction

The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement was published in 2015 to improve the reporting of multivariable prediction models.[1,2] The TRIPOD statement lists 22 items that are considered essential for informative reporting of prediction model studies. It covers diagnostic and prognostic prediction model studies and applies to all types of prediction model studies (development, external validation, and added value of predictors to existing prediction models) across all clinical domains.

In a recent study, we assessed the completeness of the reporting of 170 prediction model studies in 146 clinically diverse publications, published before the TRIPOD statement.[3] We found that prediction model studies were generally poorly reported. Items on the title and abstract were the worst affected, as they were only completely reported according to the TRIPOD reporting guideline for less than 10% of the assessed models. See supplement 1 for more details on reporting of titles and abstracts of prediction model studies.



Titles and abstracts are essential elements of a study report, as they usually are the first, and sometimes the only, part of a publication that people read. These elements facilitate the identification of potentially relevant studies by automated searches and provide the information readers need to decide whether to read the full text publication or include a study in an evidence synthesis. In many parts of the world, the abstract is often the only easily accessible part of a publication, which further emphasizes the need for including essential information in titles and abstracts. Complete reporting in conference abstracts enables better judgement of the relevance and importance of a study or presentation.

The TRIPOD statement only provides brief guidance for reporting titles and abstracts of multivariable prediction model publications. Developers of reporting guidelines addressing other study designs, like CONSORT (randomised trials), PRISMA (systematic reviews) and STARD (diagnostic test accuracy studies), have recognised similar issues and developed specific guidance for reporting abstracts for these study designs.[4-7]

Therefore, the aim of this study was to develop a list of essential items for reporting diagnostic or prognostic multivariable prediction model studies in journal and conference abstracts (TRIPOD for Abstracts), accompanied by further explanation and elaboration.

## Development of TRIPOD for Abstracts

An executive committee was formed (DGA, GSC, JBR, KGMM, LH, and PH), which established an initial list of 32 potentially relevant items for inclusion in abstracts of multivariable prediction model studies, based on the TRIPOD statement[1,2] and existing reporting guidelines for abstracts (CONSORT for Abstracts, PRISMA for Abstracts, and STARD for Abstracts)[4-7], (Supplement 2 Table 1).

The initial list was the starting point for a modified Delphi procedure undertaken as a web-based survey among the members of the original TRIPOD Group, and other clinical epidemiologists, statisticians, clinicians, and journal editors with an interest in prediction model research, who were also identified from the Cochrane Prognosis Methods Group; Grading of Recommendations Assessment, Development and Evaluation (GRADE) Prognosis Project Group; PROGnosis RESearch Strategy (PROGRESS) Partnership; and the executive committee members' personal networks. Potential panel members were invited by e-mail to participate in a web-based survey of three rounds, aiming to reach consensus on items essential to report in abstracts of prediction model studies. Supplement 2 provides details of the survey methods and results.

Of the 110 potential panel members invited to participate in the survey, 71 (65%) responded, of whom 69 completed the first round. Among the respondents were 65 (92%) clinical epidemiologists/methodologists/statisticians, 10 (14%) clinicians, and 6 (8%) journal editors (participants could be classified in more than one of these categories).

Participants were asked whether they agreed with 10 items preselected by the executive committee for inclusion in TRIPOD for Abstracts. Sixty-two (90%) agreed. They were then asked to rate to what extent they considered the remaining 22 items essential for inclusion in abstracts of prediction model studies (Supplement 2 Table 2). Participants reached consensus, defined as agreement between at least two third of the survey participants, on five of these items. Participants also had the option to provide comments and suggestions.

In the second round of the survey, the 71 first-round respondents were asked to rate the remaining 17 candidate items. Participants again had the option to provide comments and suggestions. The results of the second round are presented in Supplement 2 Table 3. Respondents (n=68; 96%) reached consensus on including another three items in TRIPOD for Abstracts.

After two rounds, the Delphi panel had agreed upon 18 items as being essential to report in abstracts of prediction model studies. Based on ratings and feedback provided in the first two rounds, the TRIPOD for Abstracts executive committee considered another three items eligible. For these items no consensus was reached, but they all scored high agreement (Supplement 2 Table 3). After discussion, the executive committee decided to add these items. After following respondent suggestions to merge some of the resulting 21 items, the draft version of TRIPOD for Abstracts consisted of 11 items (Supplement 2 Table 4).

In the third round of the survey, the panel was asked to comment on the draft version of TRIPOD for Abstracts. They were also provided with an example of complete reporting in an abstract on the development and validation of a prediction model (Supplement 2 Table 5). Of 52 (73%) respondents, 19 (37%) agreed with the draft version of TRIPOD for Abstracts without making any comments or suggestions. Thirty-three respondents (63%) provided feedback on one or more items.



This feedback was discussed during a consensus meeting with all authors. The wording of items was refined and one item on protocol availability was added to conform with other reporting guidelines for abstracts. After the consensus meeting, the final version of TRIPOD for Abstracts (12 items) was prepared for publication.

## TRIPOD for Abstracts

TRIPOD for Abstracts is a checklist of 12 items that are considered essential for inclusion in all abstracts of prediction model studies (Table 1). We developed a single checklist that can be used for all types of prediction model studies, including development, external validation, added value and model updating studies, for all types of clinical domains, for all types of predictors and outcomes and regardless the statistical approaches used (including artifical intelligence and machine learning approaches). The checklist items follow the usual structure of an abstract and are grouped under the headings Title, Background, Objectives, Methods, Results, and Discussion, with an additional item on Registration. All but one of the items overlap with items from the original TRIPOD statement (Supplement 2 Table 6). We suggest that readers consult the explanation and elaboration document that was published alongside the TRIPOD statement for detailed clarification of concepts, if needed.[2] TRIPOD for Abstracts more explicitly addresses updating of prediction models and prediction model studies using artificial intelligence or machine learning techniques than the original TRIPOD statement.

We now address the 12 TRIPOD for Abstracts items, each accompanied with an empirical example and, if needed, explanation per item. Supplement 3 provides examples of adequate reporting in abstracts of prediction model studies from varying medical disciplines and that used varying statistical approaches.

**Table 1. Essential items to include when reporting multivariable prediction model studies in journal or conference abstracts**

| Item | Description |
|---|---|
| **Title** | 1. Identification of the study as developing, validating, or updating a prediction model, the target population, and the outcome to be predicted. |
| **Background** | 2. A brief explanation of the healthcare context (including whether diagnostic or prognostic) and rationale for developing, validating, or updating the model. |
| **Objectives** | 3. Study objectives, including whether the study describes the development, validation, or updating of a model. For validation of an existing model, give the name or describe the model being validated. |
| **Methods** | 4. Study design or source of data (e.g., cohort, registry, routine care data, randomized trial), separately for the development and validation data sets, if applicable. |
| | 5. Participant eligibility criteria and setting where the data were collected. |
| | 6. Outcome to be predicted by the model, including time horizon of predictions in case of prognostic models (e.g., 3-year overall survival). |
| | 7. Statistical model or algorithm used (e.g. logistic regression, Cox regression, random forest, neural network) and approach for internal validation (for development studies). |
| **Results** | 8. Number of participants and outcome events. |
| | 9. Predictors in the final model (for development studies). |
| | 10. Performance measures, at least calibration and discrimination (with confidence intervals), and results for added value of predictors or for model-updating, if applicable. |
| **Discussion** | 11. Overall interpretation of the results, including implications for practice or research. |
| **Registration** | 12. Registration number and name of registry or repository. |

### Title

*Item 1: Identification of the study as developing, validating, or updating a prediction model, the target population, and the outcome to be predicted.*

Example: "Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease"[8]

Explanation: An informative title requires four aspects: the term *prediction model* or a synonym, the *type of prediction model study* (i.e., development, external validation,

added value, model updating, or a combination of these elements), the *target population*, and the *outcome to be predicted*.
Only 12% of the 170 reviewed prediction models described the type of prediction model study in the title (see Supplement 1).

**Background**

*Item 2: A brief explanation of the healthcare context (including whether diagnostic or prognostic) and rationale for developing, validating, or updating the model.*
Example: "Infectious endocarditis (IE) in febrile injection drug users (IDUs) is a critical diagnosis to identify in the emergency department (ED). A decision tool that identifies patients at very low risk for endocarditis using readily available clinical data could reduce admissions and cost."[9]
Explanation: An explanation of the healthcare context and rationale for the study helps abstract readers to understand the intended use of the model.



**Objectives**

*Item 3: Study objectives, including whether the study describes the development, validation, or updating of a model. For validation of an existing model, give the name or describe the model being validated.*
Example: "To evaluate the diagnostic performance of a previously derived decision instrument to rule out endocarditis in febrile IDUs (Prediction Rule for Endocarditis in Injection Drug Users [PRE-IDU]) and to develop a prediction model for likelihood of endocarditis for those who are not ruled out by PRE-IDU."[9]
Explanation: Study objectives should make clear whether the study describes the development, validation, or updating of a model. If validating an existing model, the objectives should include that model's name or description to facilitate the identification of all studies involving that model.
Study objectives were clearly reported in 76% of the publications in our review (Supplement 1).

**Methods**

*Item 4: Study design or source of data (e.g., cohort, registry, routine care data, randomized trial), separately for the development and validation data sets, if applicable.*
Example: "We performed a prospective cohort study of all trauma patients admitted to our emergency room over a 1-year period to evaluate the utility of this tool for emergency physicians to detect significant haemorrhage in the trauma patient."[10]

*Item 5: Participant eligibility criteria and setting where the data were collected.*
Example: "The Women's Health Study (WHS) is a nationwide cohort of US women free of cardiovascular disease, cancer, or other major illness at baseline from 1992 to 1995. A total of 27 542 women ages 45 to 79 years with complete ascertainment of plasma lipids and other risk factors were followed for a median of 10 years."[11]

*Item 6: Outcome to be predicted by the model, including time horizon of predictions in case of prognostic models (e.g., 3-year overall survival).*
Example: "The outcome was 5-year all-cause mortality..."[12]

Explanation: Including the study design and data source (item 4), participant eligibility criteria and setting (item 5), and outcome to be predicted (item 6) provides insight into the prediction model's applicability and generalizability. Describing the data source also helps the reader to judge the risk of bias, which varies with study design.[13,14] Also, the predictive ability of a model is very dependent on the predicted outcome and on the prediction horizon.
The setting was reported in 69% of the 170 reviewed models, study design or data source in 76%, study participants in 78%, and predicted outcomes in 95% (Supplement 1).

*Item 7: Statistical model or algorithm used (e.g., logistic regression, Cox regression, random forest, neural network), and approach for internal validation (for development studies).*
Example: "In this retrospective cohort study, 6-month, 1-year, and 2-year mortality prediction models with recurrent neural networks used patient demographic information and topics generated from clinical notes within Partners HealthCare System, ... The models were trained using a data set of 24 229 patients and validated using another data set of 2692 patients."[15]
Example: "Prognostic models were developed using proportional odds ordinal logistic regression using patient characteristics and baseline and 3-month patient reported outcome scores. Models were fit for each outcome stratified by type of surgical procedure. ... Models were internally validated using bootstrap resampling."[16]
Explanation: The full text of a prediction model study publication should contain enough details about the statistical model to understand and verify the approach taken. In contrast, the abstract should just make clear what statistical model or algorithm was applied and, for model development and updating, the approach for internal validation (item 7). Internal validation is important for assessing overfitting of the developed or updated model and adjusting for optimism in model performance.[17-19] Reporting this essential step in model development or updating in the abstract helps the reader to judge the study's risk of bias.[13,14]

The statistical methods used for model development or validation were reported for about half (53%) of the models in the review (Supplement 1).

**Results**

*Item 8: Number of participants and outcome events.*

Example: "The derivation and validation cohort consisted of 240 and 793 patients with COPD, of whom 29% and 28%, respectively, experienced an exacerbation during follow-up."[8]

Explanation: The number of participants and outcome events are important for interpreting a prediction model's precision and the risk of bias in its performance estimates.[13,14] The lower the sample size and particularly the lower the number of study participants with the outcome, the higher the risk of bias in the estimates of a model's predictive performance measures.

Overall sample size and number of participants with the outcome were reported for 94% and 49% of the 170 assessed models, respectively (Supplement 1).



*Item 9: Predictors in the final model (for development studies).*

Example: "The final model included four easily assessable variables: exacerbations in the previous year, pack years of smoking, level of obstruction, and history of vascular disease, ..."[8]

Explanation: For development studies, the abstract should report which predictors were included in the final model. If there are too many predictors to list in an abstract, authors can instead describe predictor categories (e.g., socio-demographical predictors, history taking and physical exam items, laboratory or imaging tests, disease characteristics).

Predictors in the final model were reported for 63% of the model development studies in our review (Supplement 1).

*Item 10: Performance measures, at least calibration and discrimination (with confidence intervals), and results for added value of predictors or for model-updating, if applicable.*

Example: " The ADO score was discriminatory for predicting 3-year mortality (AUC= 0.74; 95% CI: 0.69-0.79). Similar performance was found for 1- (AUC= 0.73; 0.66-0.80) and 2-year mortality (0.72; 0.67-0.76). The ADO score showed reasonable calibration for predicting 3-year mortality (calibration slope 0.95; 0.70-1.19) but over-predicted in cases with higher predicted risks of mortality at 1 (0.79; 0.45-1.13) and 2-year (0.79; 0.57-1.01) mortality."[20]

Explanation: The abstract for a prediction model study should include model performance results (item 10). At least calibration and discrimination (with confidence

intervals) should be presented (preferably the optimism-corrected performance measures), as these are the two key aspects for characterizing prediction model performance. The results of the added value of predictors and model updates (e.g. increase in c-statistic of the model after adding predictors or updating the model) should be reported, if this was undertaken. Some measures, like calibration, are often presented graphically, however they can be quantified, e.g. calibration slope or calibration in the large. We suggest that authors preferably report these quantitative calibration measures in the abstract. If allowed, such as in conference abstracts, a graph could also be included.
Discrimination performance measures were reported more often (44% of the 170 models) than calibration measures (11%, Supplement 1).

**Discussion**

*Item 11: Overall interpretation of the results, including implications for practice or research.*
Example: "The pooled cohort risk score appears to overestimate CV risk but this apparent over-prediction could be a result of treatment. In the absence of a validated score in an untreated population, the pooled cohort risk score appears to be appropriate for use in a primary care setting."[21]
Explanation: A brief concluding statement of the overall interpretation of the results, including main limitations and implications for clinical practice or research (item 11) enables readers to consider how the results apply to them.
Main conclusions were reported in 91% of the 170 assessed models (Supplement 1).

**Registration**

*Item 12: Registration number and name of registry or repository*
Example: " ... a large prospective cohort study (PREP-946) for development of prognostic models… TRIAL REGISTRATION: ISRCTN40384046."[22]
Example: "We developed a simple/practical scoring rule (logistic regression model) for recurrent CDI using data from 2 large phase 3 clinical trials. ... CLINICAL TRIALS REGISTRATION: NCT00314951 and NCT00468728."[23]
Explanation: Although registration of prediction model studies is not yet common practice, it is helpful to indicate the availability of a study protocol or data in a register or repository, and provide relevant registration numbers for abstract readers. The first example above reflects the reporting of a registered prediction model study. In the second example the authors refer to two registered randomized trials of which data were used to develop a prognostic prediction model for the risk of recurrence of Clostridium difficile infection (CDI) in patients recently diagnosed with CDI.

## Discussion

Although abstracts cannot and should not replace full research reports in the communication of research findings, they have an important role in informing readers what was done. TRIPOD for Abstracts contains items that are considered essential for inclusion in all abstracts of prediction model studies. This checklist is applicable to any type of prediction model study, regardless whether it addresses a diagnostic or prognostic model; the development, validation or updating of a model, or estimating the added value of one or more predictors to an existing model; and whether prevailing or modern statistical or machine learning techniques are used. Although the checklist presents the items in the typical order of an abstract, the items do not have to appear in abstracts in this strict order. How the items are incorporated into the abstract will depend on journal and conference requirements. These items should also be seen as the minimum set of information that is required for informative abstracts on prediction models.



During the development of TRIPOD for Abstracts, several survey respondents expressed concerns about the limited space typically allowed for abstracts. Although challenging, we believe it is possible to provide all of the essential information needed for a prediction model study within 250 to 350 words, as shown by examples of adequate reporting provided in Supplement 3.

Without complete reporting of a study, the efforts spent in conducting the research can be considered wasted.[24] This includes the reporting titles and abstracts. Reporting guidelines are tools primarily targeted at researchers to enhance the transparency and completeness of the reporting of their research. However, peer reviewers and journal editors can also use these guidelines to check reporting completeness and prevent the publication of poorly reported research.

We developed this extension of the TRIPOD statement to improve the reporting of prediction model studies in abstracts. Comparable initiatives developing reporting guidelines for abstracts for other study designs have been evaluated in systematic reviews that compare reporting in abstracts before and after the publication of these guidelines. These evaluations have found more complete reporting in abstracts after the introduction of guidelines for abstracts, although all have concluded that there is still room for improvement.[25-30]

TRIPOD for Abstracts will contribute to improved reporting in abstracts of prediction model studies. Readers and reviewers will be better supported in identifying

potentially relevant prediction model studies and assessing the applicability and validity of the findings from abstracts, thus ensuring they can take full advantage of the available evidence from this type of studies.

## Acknowledgements

We are grateful to all who contributed to the development of TRIPOD for Abstracts by responding to our Delphi survey. In addition, we would like to thank Romin Pajouheshnia (Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University) for contributing to the assessment of the reporting of prediction model studies in titles and abstracts. Finally, we would like to thank Dr Jennifer A. de Beyer (Centre for Statistics in Medicine, University of Oxford) for English language editing.

## References


1. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55-63.
2. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1-73.
3. Heus P, Damen J, Pajouheshnia R, Scholten R, Reitsma JB, et al. Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. *BMC Med.* 2018;16(1):120.
4. Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, et al. PRISMA for Abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med.* 2013;10(4):e1001419.
5. Cohen JF, Korevaar DA, Gatsonis CA, Glasziou PP, Hooft L, et al. STARD for Abstracts: essential items for reporting diagnostic accuracy studies in journal or conference abstracts. *BMJ.* 2017;358:j3751.
6. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet.* 2008;371(9609):281-283.
7. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med.* 2008;5(1):e20.
8. Bertens LC, Reitsma JB, Moons KG, van Mourik Y, Lammers JW, et al. Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis.* 2013;8:493-499.
9. Chung-Esaki H, Rodriguez RM, Alter H, Cisse B. Validation of a prediction rule for endocarditis in febrile injection drug users. *Am J Emerg Med.* 2014;32(5):412-416.
10. Lawton LD, Roncal S, Leonard E, Stack A, Dinh MM, et al. The utility of Advanced Trauma Life Support (ATLS) clinical shock grading in assessment of trauma. *Emerg Med J.* 2014;31(5):384-389.
11. Cook NR, Ridker PM. Further insight into the cardiovascular risk calculator: the roles of statins, revascularizations, and underascertainment in the Women's Health Study. *JAMA Intern Med.* 2014;174(12):1964-1971.
12. Martinez-Garcia MA, de Gracia J, Vendrell Relat M, Giron RM, Maiz Carro L, et al. Multidimensional approach to non-cystic fibrosis bronchiectasis: the FACED score. *Eur Respir J.* 2014;43(5):1357-1367.
13. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med.* 2019;170(1):51-58.

14. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med*. 2019;170(1):W1-w33.
15. Wang L, Sha L, Lakin JR, Bynum J, Bates DW, et al. Development and Validation of a Deep Learning Algorithm for Mortality Prediction in Selecting Patients With Dementia for Earlier Palliative Care Interventions. *JAMA Netw Open*. 2019;2(7):e196972.
16. Rundell SD, Pennings JS, Nian H, Harrell FE, Jr., Khan I, et al. Adding 3-month patient data improves prognostic models of 12-month disability, pain, and satisfaction after specific lumbar spine surgical procedures: development and validation of a prediction model. *Spine J*. 2019.
17. Harrell FE, Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag; 2001.
18. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York: Springer-Verlag; 2009.
19. Riley RD, van der Windt D, Croft P, Moons KGM. *Prognosis Research in Healthcare: Concepts, Methods, and Impact*. Oxford: Oxford University Press; 2019.
20. Keene SJ, Jordan RE, Franssen FM, de Vries F, Martin J, et al. External Validation Of The Updated ADO Score In COPD Patients From The Birmingham COPD Cohort. *Int J Chron Obstruct Pulmon Dis*. 2019;14:2395-2407.
21. Chia YC, Lim HM, Ching SM. Validation of the pooled cohort risk score in an Asian population - a retrospective cohort study. *BMC Cardiovasc Disord*. 2014;14:163.
22. Thangaratinam S, Allotey J, Marlin N, Dodds J, Cheong-See F, et al. Prediction of complications in early-onset pre-eclampsia (PREP): development and external multinational validation of prognostic models. *BMC Med*. 2017;15(1):68.
23. D'Agostino RB, Sr., Collins SH, Pencina KM, Kean Y, Gorbach S. Risk estimation for recurrent Clostridium difficile infection based on clinical factors. *Clin Infect Dis*. 2014;58(10):1386-1393.
24. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267-276.
25. Chen J, Li Z, Liu B, Gan X, Li C, Yu H. Quality improvement in randomized controlled trial abstracts in prosthodontics since the publication of CONSORT guideline for abstracts: a systematic review. *J Dent*. 2018:23-29.
26. Chhapola V, Tiwari S, Brar R, Kanwal SK. Reporting quality of trial abstracts-improved yet suboptimal: A systematic review and meta-analysis. *J Evid Based Med*. 2018;11(2):89-94.

27. Ghimire S, Kyung E, Lee H, Kim E. Oncology trial abstracts showed suboptimal improvement in reporting: a comparative before-and-after evaluation using CONSORT for Abstract guidelines. *J Clin Epidemiol*. 2014;67(6):658-666.
28. Mbuagbaw L, Thabane M, Vanniyasingam T, Borg Debono V, Kosa S, et al. Improvement in the quality of abstracts in major clinical journals since CONSORT extension for abstracts: a systematic review. *Contemp Clin Trials*. 2014;38(2):245-250.
29. Speich B, Mc Cord KA, Agarwal A, Gloy V, Gryaznov D, et al. Reporting Quality of Journal Abstracts for Surgical Randomized Controlled Trials Before and After the Implementation of the CONSORT Extension for Abstracts. *World J Surg*. 2019:2371-2378.
30. Vasquez-Cardenas J, Zapata-Norena O, Carvajal-Florez A, Barbosa-Liz DM, Giannakopoulos NN, Faggion CM, Jr. Systematic reviews in orthodontics: Impact of the PRISMA for Abstracts checklist on completeness of reporting. *Am J Orthod Dentofacial Orthop*. 2019;156(4):442-452.e412.

## Supplementary material



### Supplement 1 – Reporting of prediction model studies in abstracts

To examine the reporting of prediction model studies in abstracts, we used the set of publications previously identified for the baseline measurement of adherence to TRIPOD.[1] This set consists of 146 clinically diverse (n=122; 84%) publications, from which the reporting of 170 models was assessed, 73 (43%) concerning model development, 43 (25%) external validation, 33 (19%) incremental value, and 21 (12%) combined development and external validation of the same model. Further details regarding the set, including the methods of collecting the publications and assessing the reporting of the included models can be found elsewhere.[1,2]

The table below shows the assessment of the completeness of reporting of the TRIPOD title and abstract elements in the 170 included models. We found that the reporting of titles and abstracts was incomplete, with full adherence to the TRIPOD reporting guideline in less than 10% of the 170 models.

**Results of the assessment of the reporting of 170 prediction model studies in titles and abstracts of 146 publications.[1]**

| | Development | External validation | Incremental value | Development and external validation | Overall |
|---|---|---|---|---|---|
| | N=73 | N=43 | N=33 | N=21 | N=170 |
| **Title** | | | | | |
| **Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.** ***[adhered to if all elements below were scored as "yes"]*** | **1 (1%)** | **4 (9%)** | **3 (9%)** | **0 (0%)** | **8 (5%)** |
| The words developing/development, validation/validating, incremental/added value (or synonyms) are reported in the title | 5 (7%) | 8 (19%) | 5 (15%) | 3 (14%) | 21 (12%) |
| The words prediction, risk prediction, prediction model, risk models, prognostic models, prognostic indices, risk scores (or synonyms) are reported in the title | 57 (78%) | 34 (79%) | 27 (82%) | 16 (76%) | 134 (79%) |
| The target population is reported in the title | 53 (73%) | 33 (77%) | 24 (73%) | 15 (71%) | 125 (74%) |
| The outcome to be predicted is reported in the title | 54 (74%) | 28 (65%) | 27 (82%) | 13 (62%) | 122 (72%) |
| **Abstract** | | | | | |
| **Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.** ***[adhered to if all elements below were scored as "yes" or "not applicable"]*** | **1 (1%)** | **2 (5%)** | **0 (0%)** | **0 (0%)** | **3 (2%)** |
| The objectives are reported in the abstract | 62 (85%) | 33 (77%) | 22 (67%) | 13 (62%) | 130 (76%) |
| Sources of data are reported in the abstract *E.g. Prospective cohort, registry data, RCT data.* | 53 (73%) | 35 (81%) | 25 (76%) | 16 (76%) | 129 (76%) |
| The setting is reported in the abstract *E.g. Primary care, secondary care, general population, adult care, or paediatric care. The setting should be reported for both the development and validation datasets, if applicable.* | 51 (70%) | 35 (81%) | 20 (61%) | 12 (57%) | 118 (69%) |

| | | | | | |
|---|---|---|---|---|---|
| A general definition of the study participants is reported in the abstract<br>*E.g. patients with suspicion of certain disease, patients with a specific disease, or general eligibility criteria.* | 60 (82%) | 37 (86%) | 23 (70%) | 13 (62%) | 133 (78%) |
| The overall sample size is reported in the abstract | 71 (97%) | 39 (91%) | 31 (94%) | 18 (86%) | 159 (94%) |
| The number of events (or % outcome together with overall sample size) is reported in the abstract<br>*If a continuous outcome was studied, score Not applicable* | 39/69 (57%)* | 21/42 (50%)* | 17/32 (53%)* | 3/19 (16%)* | 80/162 (49%)* |
| Predictors included in the final model are reported in the abstract. For validation studies of well-known models, at least the name/acronym of the validated model is reported<br>*Broad descriptions are sufficient, e.g. 'all information from patient history and physical examination'. Check in the main text whether all predictors of the final model are indeed reported in the abstract.* | 46 (63%) | 22 (51%) | 13 (39%) | 20 (95%) | 101 (59%) |
| The outcome is reported in the abstract | 71 (97%) | 42 (98%) | 30 (91%) | 19 (90%) | 162 (95%) |
| Statistical methods are described in the abstract<br>*For model development, at least the type of statistical model should be reported. For validation studies a quote like "model's discrimination and calibration was assessed" is considered adequate. If done, methods of updating should be reported.* | 47 (64%) | 20 (48%) | 12 (36%) | 11 (52%) | 90 (53%) |
| Results for model discrimination are reported in the abstract<br>*This should be reported separately for development and validation if a study includes both development and validation.* | 34 (47%) | 20 (47%) | 13 (39%) | 8 (38%) | 75 (44%) |
| Results for model calibration are reported in the abstract<br>*This should be reported separately for development and validation if a study includes both development and validation.* | 7 (10%) | 9 (21%) | 1 (3%) | 1 (5%) | 18 (11%) |
| Conclusions are reported in the abstract<br>*In publications addressing both model development and validation, there is no need for separate conclusions for both; one conclusion is sufficient.* | 68 (93%) | 35 (81%) | 31 (94%) | 20 (95%) | 154 (91%) |

n (%)

*Percentages are based on number of models for which that item was applicable (and should have been reported)

**Supplement 2 – Survey methods and results**

For the development of TRIPOD for Abstracts, a list of essential items for informative reporting of diagnostic or prognostic multivariable prediction model studies in both journal and conference abstracts, we sought input of and consensus among experts in the field of prediction modelling using a modified Delphi procedure.[3,4] We planned a web-based survey of a maximum of three rounds using SurveyMonkey, an online software tool to develop and run surveys.[5]

Here, we provide a detailed summary of the methods and results of this survey.

*Survey items*

Based on the TRIPOD Statement,[6,7] as well as on previous initiatives of reporting guidelines for abstracts (i.e. CONSORT for Abstracts, STARD for Abstracts, and PRISMA for Abstracts),[8-10] the TRIPOD for Abstracts executive committee (DGA, GSC, JBR, KGMM, LH, and PH) created an initial list of 32 potentially relevant items to report in abstracts of prediction model studies (Supplemental Table 1). This initial item list was the starting point for the first round of the survey. Items submitted to the Delphi panel in subsequent rounds of the survey depended on the results of the preceding round (see below). The first author (PH) drafted each of the survey rounds in SurveyMonkey. Before asking the Delphi panel to participate, a survey round was tested by at least two other authors (JAAGD, JBR, LH, KGGM).

*Survey participants*

We invited the members of the TRIPOD Group to participate in the survey. In addition, we approached other clinical epidemiologists, statisticians, clinicians, and journal editors with an interest in prediction model research, who were identified from the Cochrane Prognosis Methods Group; GRADE (Grading of Recommendations Assessment, Development and Evaluation) Prognosis Project Group, PROGRESS (PROGnosis RESearch Strategy) Partnership, or from the personal networks of the executive committee members.

*Survey administration*

Potential panel members received the invitation to participate in the survey by e-mail. This e-mail was sent by one of the members of the TRIPOD for Abstracts executive committee (KGMM). It explained the overarching aim of the project (i.e. development of TRIPOD for Abstracts) and the aim of the survey (i.e. to identify and reach consensus on items essential to report in abstracts of prediction model studies). The e-mail contained a web link to the first round of survey with an estimated completion time of 10 minutes.

*Round 1*

First, participants were asked whether they agreed with 10 items preselected by the executive committee to be definitely included in TRIPOD for Abstracts. If case of disagreement, they were asked to indicate which item(s) should be considered by the Delphi panel. Then, we asked them to rate the remaining 22 items on a five-point scale. A rating of "1" meant that the item should certainly not be included in TRIPOD for Abstracts and a rating of "5" meant that the item should certainly be included, as it is essential to report in all abstracts of prediction model studies. The 22 items were structured under the headings 'Rationale / Background', 'Methods', 'Results', and 'Discussion/Conclusion'. After each heading with corresponding items participants had the option to provide comments and suggestions. For the rating of the 22 items, we grouped scores in three categories: low (rates 1 and 2; item should not be included in TRIPOD for Abstracts), middle (rate 3; inconclusive whether item should be included, and high rates (rates 4 and 5; item should be included). Consensus was considered as reached if ≥2/3 (67%) of survey participants rated an item in either the high (4-5) or low category (1-2). In all other cases consensus was not considered as reached.

Invitations to participate were sent on April 1, 2016, followed by two reminders after 10 days each, that were sent to the people that had not responded up to then. The survey was closed on May 26, 2016, after one final call a week before.

Of the 110 potential panel members invited, 71 (65%) responded. Among the respondents were 65 (92%) clinical epidemiologists/methodologists/statisticians, 10 (14%) clinicians, and 6 (8%) journal editors (numbers add up to over 71, as people could have been classified to more than one category). Sixty-nine of the respondents completed the questionnaire. Of these 69 participants, 62 (90%) agreed that the 10 preselected items are essential to report in abstracts of prediction model studies and should be included in TRIPOD for Abstracts. In addition, consensus was reached for inclusion of five of the 22 items that were rated on a five point scale by the participants (Supplemental Table 2). We deduced the following themes from the comments and suggestions provided by the participants. Several participants raised that developing a single reporting guideline for abstracts of prediction models is challenging, as essential information to report is strongly related to the objective of a prediction model study (i.e., development, external validation, incremental value assessment) and the (clinical) context. In addition, several participants expressed their concerns with regard to the feasibility of reporting all essential information in relation to word limits set by medical journals. Furthermore, it turned out that it was insufficiently clear to participants what was meant by the items 'prediction horizon' and 'risk groups'. In

the second round of the survey we submitted these two items to the Delphi panel with additional explanation. Also, based on comments provided, we slightly adapted the wording of the items 'Study location', 'Internal validation technique', 'Blinding of outcome assessment' and 'Blinding of predictor assessment'.

*Round 2*

Respondents to the first round of the survey received a summary of the results, including the list of 15 items on which consensus was reached, and were invited to participate in the second round. In this second round they were asked to rate on a three-point scale (no / no opinion / yes) whether they considered the remaining 17 candidate items essential to report in (nearly) all abstracts of prediction model studies. The items were presented to the panel as a list with the option to tick the answer of preference for each item, followed by a comment box for any comments or suggestions. Again, consensus was defined as agreement between at least two third of the survey participants, i.e. when 67% or more of the survey participants rated the item as either "yes" or "no".



Invitations to participate in the second round of the survey were sent on July 25, 2016. Two reminders (after three and two weeks, respectively) were sent to participants that had not responded up to then. We closed the survey on September 12, 2016.

Of the 71 persons invited to participate, 68 (96%) responded. Respondents reached consensus on including another three items in TRIPOD for Abstracts (Supplemental Table 3). Twenty-six participants provided a comment or suggestion after rating the items. Again, concerns were expressed regarding the limited space there usually is in abstracts. Also some participants noted that what is considered essential is strongly related to the objective (type) of a prediction model study. In addition, helpful suggestions were provided to combine several of the items and to include an item on the availability of a protocol or registration number, conform other reporting guidelines for abstracts.

After this second round of the survey the Delphi panel already agreed upon 18 items as being essential to report in abstracts of prediction model studies. Based on the rating scores in round one and two of the survey the TRIPOD for Abstracts executive committee decided that another round of asking feedback on whether items should be included in TRIPOD for Abstracts would not be necessary. For three of the remaining 14 items no consensus was reached, but they all scored high agreement (Supplemental Table 3). After discussion, the committee decided to add these items as well, which resulted in a list of 21 items. Following panel members' suggestions to merge some

of the items, the committee reduced the list of 21 items, resulting in a draft version of TRIPOD for Abstracts consisting of 11 items (Supplemental Table 4).

*Round 3*

In the third round of the survey the Delphi-panel was asked to comment on the draft version of TRIPOD for Abstracts (Supplemental Table 4). In addition, panel members were provided with an example of complete reporting in an abstract on development and validation of a prognostic prediction model (Supplemental Table 5). Subsequently, panel members were asked whether they had any comments or suggestions regarding the draft version of TRIPOD for Abstracts. If they did not, the survey ended. If they did have comments or suggestions, they were linked to a next page where they could provide their feedback on each of the items. It was stressed that at this stage the question was not whether items should be included (although any major concerns could be shared), but that we would like to receive suggestions for improvement of the wording of items, in order to make them as clear and unambiguous as possible. The survey ended with a comment box for any remaining overall comments.

Invitations to participate in the third round of the survey were sent on April 16, 2019. After two weeks, a reminder was sent to participants that had not responded up to then. We closed the survey on May 22, 2019.

Of the 71 original Delphi panel members, one had deceased and for another we could not track down a valid e-mail address. The two authors not being part of the TRIPOD for Abstracts executive committee (JAAGD and RS) were also invited to participate in this third round of the survey. Of the 71 persons invited, 52 (73%) responded, of which one stated not to participate because of a potential conflict of interest, and another only filled in a name and did not answer any question. Nineteen respondents agreed with the draft version of TRIPOD for Abstracts without any comments or suggestions. Thirty-three respondents provided feedback regarding one or more items.

*Finalizing TRIPOD for Abstracts*

The first author (PH) prepared a final consensus meeting with all authors, in which the feedback on the draft version of TRIPOD for Abstracts provided in the third round of the survey was discussed. This led to textual adjustments of some items. In addition, based on the feedback provided, we decided to add an item regarding the availability of a protocol, registration number or repository (machine learning), conform other reporting guidelines for abstracts, which resulted in a final list of 12 items. After the consensus meeting the final version of TRIPOD for Abstracts was drafted.

**Supplemental Table 1. Initial list of 32 potentially relevant items to report in abstracts of prediction model studies**

| | | |
|---|---|---|
| **Title, background and objectives** | | |
| * | 1. | Title<br>*Identify the study as developing and/or validating a multivariable prediction model, the target population, the (main) outcome to be predicted.* |
| | 2. | Rationale / background |
| * | 3. | Objectives<br>*Specify the objectives, including whether the study describes the development or validation of the model or both.* |
| **Methods** | | |
| | 4. | Source of data<br>*E.g. prospective cohort, registry data, RCT etc.* |
| * | 5. | Main eligibility criteria |
| * | 6. | Setting<br>*The setting should be reported for both the development and validation datasets, if applicable.* |
| | 7. | Key study dates<br>*Including start of accrual, end of accrual, and if applicable end of follow-up.* |
| | 8. | Number of centers |
| | 9. | Study location<br>*E.g. country.* |
| * | 10. | Outcome |
| | 11. | Prediction horizon |
| | 12. | Type of statistical model used |
| | 13. | Internal validation done |
| | 14. | Internal validation technique |
| | 15. | Blinding outcome assessment |
| | 16. | Blinding predictor assessment |
| | 17. | Risk groups |
| **Results** | | |
| * | 18. | Sample size |
| | 19. | Relevant baseline characteristics of patients |
| * | 20. | Predictors included in the final model<br>*For validation studies of well-known models, at least the name/acronym of the validated model is reported.* |
| * | 21. | Number of events (or % outcome together with overall sample size) |
| * | 22. | Results for discrimination<br>*Should be reported separately for development and validation if a study includes both development and validation.* |
| | 23. | Confidence intervals (or standard error) around estimates for discrimination |

| | |
|---|---|
| | 24. Results for calibration<br>*Should be reported separately for development and validation if a study includes both development and validation.* |
| | 25. Regression coefficients of the final model (model development studies) |
| | 26. Confidence intervals (or standard error) for regression coefficients |
| | 27. Results of model updating / recalibration |
| | 28. Results for added value |
| **Discussion and conclusion** | |
| | 29. Potential clinical use / implications for practice or future research |
| | 30. Limitations |
| * | 31. Conclusions |
| | 32. Sources of funding |

* Item considered to be essential in abstracts of prediction model studies by the TRIPOD for Abstracts executive committee and therefore suggested to be definitely included in TRIPOD for Abstracts.

**Supplemental Table 2. First round of the survey: results for the rating of 22 potentially relevant items to report in abstracts of prediction model studies by the panel (n=69)**

| Items | Mean score | Distribution of ratings | | | Consensus* |
|---|---|---|---|---|---|
| | | 1+2 | 3 | 4+5 | |
| 1. Rationale / background | 3.8 | 19% | 18% | 63% | No |
| 2. Source of data<br>*E.g. prospective cohort, registry data, RCT etc.* | 4.4 | 4.4% | 8.8% | **87%** | **Yes** |
| 3. Key study dates<br>*Including start of accrual, end of accrual, and if applicable end of follow-up* | 2.9 | 34% | 37% | 29% | No |
| 4. Number of centers | 2.8 | 41% | 32% | 27% | No |
| 5. Study location<br>*E.g. country* | 3.4 | 28% | 19% | 53% | No |
| 6. Prediction horizon | 3.8 | 19% | 15% | 66% | No |
| 7. Type of statistical model used | 3.8 | 10% | 27% | 63% | No |
| 8. Internal validation done | 3.8 | 13% | 19% | **68%** | **Yes** |
| 9. Internal validation technique | 3.1 | 34% | 34% | 32% | No |
| 10. Blinding of outcome assessment | 3.0 | 37% | 29% | 34% | No |
| 11. Blinding of predictor assessment | 3.0 | 38% | 29% | 32% | No |
| 12. Risk groups | 3.1 | 32% | 31% | 37% | No |
| 13. Relevant baseline characteristics of patients | 3.2 | 29% | 32% | 38% | No |
| 14. Confidence intervals (or standard error) around estimates for discrimination | 4.1 | 10% | 18% | **72%** | **Yes** |
| 15. Results for calibration<br>*Should be reported separately for development and validation if a study includes both development and validation* | 3.8 | 12% | 19% | **69%** | **Yes** |
| 16. Regression coefficients of the final model (model development studies) | 2.4 | 59% | 25% | 16% | No |
| 17. Confidence intervals (or standard error) for regression coefficients | 2.2 | 66% | 19% | 15% | No |
| 18. Results of model updating / recalibration | 3.5 | 27% | 18% | 56% | No |
| 19. Results for added value | 3.7 | 15% | 25% | 60% | No |
| 20. Potential clinical use / implications for practice or future research | 4.2 | 9% | 10% | **81%** | **Yes** |
| 21. Limitations | 3.3 | 25% | 34% | 41% | No |
| 22. Sources of funding | 2.8 | 49% | 24% | 28% | No |

*Consensus was considered as reached if ≥2/3 (66.7%) of survey participants rated the item in either the high (4-5) or low category (1-2). In all other cases consensus was not reached.

**Supplemental Table 3. Second round of the survey: results for the rating of 17 items remaining from round 1 to report in abstracts of prediction model studies by the panel (n=68)**

| Items | Mean score | Distribution of ratings | | | Consensus* |
|---|---|---|---|---|---|
| | | 1 No | 2 No opinion | 3 Yes | |
| 1. Rationale / background | 2.6 | 18% | 6% | **76%** | **Yes** |
| 2. Key study dates<br>*Including start of accrual, end of accrual, and, if applicable, end of follow-up.* | 1.9 | 49% | 18% | 34%† | No |
| 3. Number of centers | 1.9 | 46% | 22% | 32% | No |
| 4. Study location<br>*E.g. geographical region, country.* | 2.5 | 19% | 12% | **69%** | **Yes** |
| 5. Prediction horizon<br>*Time frame in prognostic studies, e.g. 10-year risk* | 2.9 | 3% | 4% | **93%** | **Yes** |
| 6. Type of statistical model used | 2.5 | 19% | 15% | 66% | No‡ |
| 7. Internal validation technique<br>*(the focus is on the actual internal validation technique that was used; whether internal validation is done, was already included as an essential item in the first round of the survey)* | 1.9 | 50% | 15% | 35% | No |
| 8. Blinding of outcome assessment<br>*Outcome assessed without knowledge of predictors, if applicable (subjective outcomes).* | 1.9 | 47% | 19% | 34% | No |
| 9. Blinding of predictor assessment<br>*For the outcome and other predictors.* | 1.8 | 51% | 22% | 26%† | No |
| 10. Risk groups<br>*Stated which risk groups were created, if applicable.* | 2.2 | 28% | 29% | 43% | No |
| 11. Relevant baseline characteristics of patients | 2.1 | 35% | 22% | 43% | No |
| 12. Regression coefficients of the final model (model development studies) | 1.8 | 54% | 16% | 29%† | No |
| 13. Confidence intervals (or standard error) for regression coefficients | 1.7 | 57% | 19% | 24% | No |
| 14. Results of model updating / recalibration, if applicable | 2.3 | 22% | 26% | 51%† | No‡ |
| 15. Results for added value of predictors, if applicable | 2.5 | 12% | 26% | 62% | No‡ |
| 16. Limitations | 2.0 | 35% | 28% | 37% | No |
| 17. Sources of funding | 1.7 | 54% | 26% | 19%† | No |

*Consensus was considered as reached if ≥2/3 (67%) of survey participants rated the item in either the yes or no category.

†Percentages do not add up to 100% due to rounding.

‡Item that eventually was included in TRIPOD for Abstracts after discussion within the TRIPOD for Abstracts executive committee.

**Supplemental Table 4. Draft version of TRIPOD for Abstracts (11 items) that was submitted to the panel in the third round of the survey**

| Item | Description |
|---|---|
| **Title** | 1. Identification of the study as developing and/or validating a prediction model, the target population, and the outcome to be predicted. |
| **Background** | 2. A brief explanation of medical context (including whether diagnostic or prognostic) and rationale. |
| **Objectives** | 3. Study objectives, including whether the study describes the development or validation of a model or both. For validation of an existing model, describe the name of the model that is being validated. |
| **Methods** | 4. Study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable. |
| | 5. Eligibility criteria for participants and settings where the data were collected, including geographical location. |
| | 6. Outcome(s) to be predicted by the model, including time frame in prognostic model (e.g., 10-year risk). |
| | 7. Use of regression (logistic/survival) or non-regression based statistical model and whether internal validation was done. |
| **Results** | 8. Number of participants and outcome events. |
| | 9. Predictors in the final model (for development studies only). |
| | 10. Results for discrimination (with confidence intervals) and calibration; and results for added value of predictors and/or model-updating (if applicable). |
| **Discussion** | 11. Overall interpretation of the results, including the potential clinical use of the model and implications for future research. |

**Supplemental Table 5. Example of complete reporting in an abstract on development and validation of a prognostic prediction model that was provided to the panel in the third round of the survey**

**Development and validation of a model to predict the 2-year risk of exacerbations in chronic obstructive pulmonary disease.**
*(word count n=274)*

PURPOSE:
Prognostic models for exacerbations in patients with chronic obstructive pulmonary disease (COPD) are scarce. Our aim was to develop and validate a new model to predict exacerbations within two years in patients with COPD.

PATIENTS AND METHODS:
The derivation cohort consisted of Dutch patients aged 65 years or over with a COPD diagnosis, who were followed up over 24 months. The external validation cohort consisted of another Dutch cohort of COPD patients, aged 50 years or over. Exacerbations of COPD were defined as symptomatic deterioration requiring pulsed oral steroid use or hospitalization. Logistic regression analysis including backward selection and shrinkage (determined with bootstrapping) were used to develop the final model and to adjust for overfitting. The adjusted regression coefficients were applied in the validation cohort to assess calibration of the predictions and calculate changes in discrimination applying C-statistics.

RESULTS:
The derivation and validation cohort consisted of 240 and 793 patients with COPD, of whom 29% and 28%, respectively, experienced an exacerbation during follow-up. The final model included four easily assessable variables: exacerbations in the previous year, pack years of smoking, level of obstruction, and history of vascular disease, with a C-statistic of 0.75 (95% confidence interval [CI]: 0.69-0.82). Predictions were well calibrated in the validation cohort, with a small loss in discrimination potential (C-statistic 0.66 [95% CI 0.61-0.71]).

CONCLUSION:
Our newly developed prediction model can help clinicians to predict the risk of future exacerbations in individual patients with COPD, including those with mild disease. An implementation study should be performed to determine the impact of our prediction model on daily practice in terms of patient outcome and the use of health care resources.

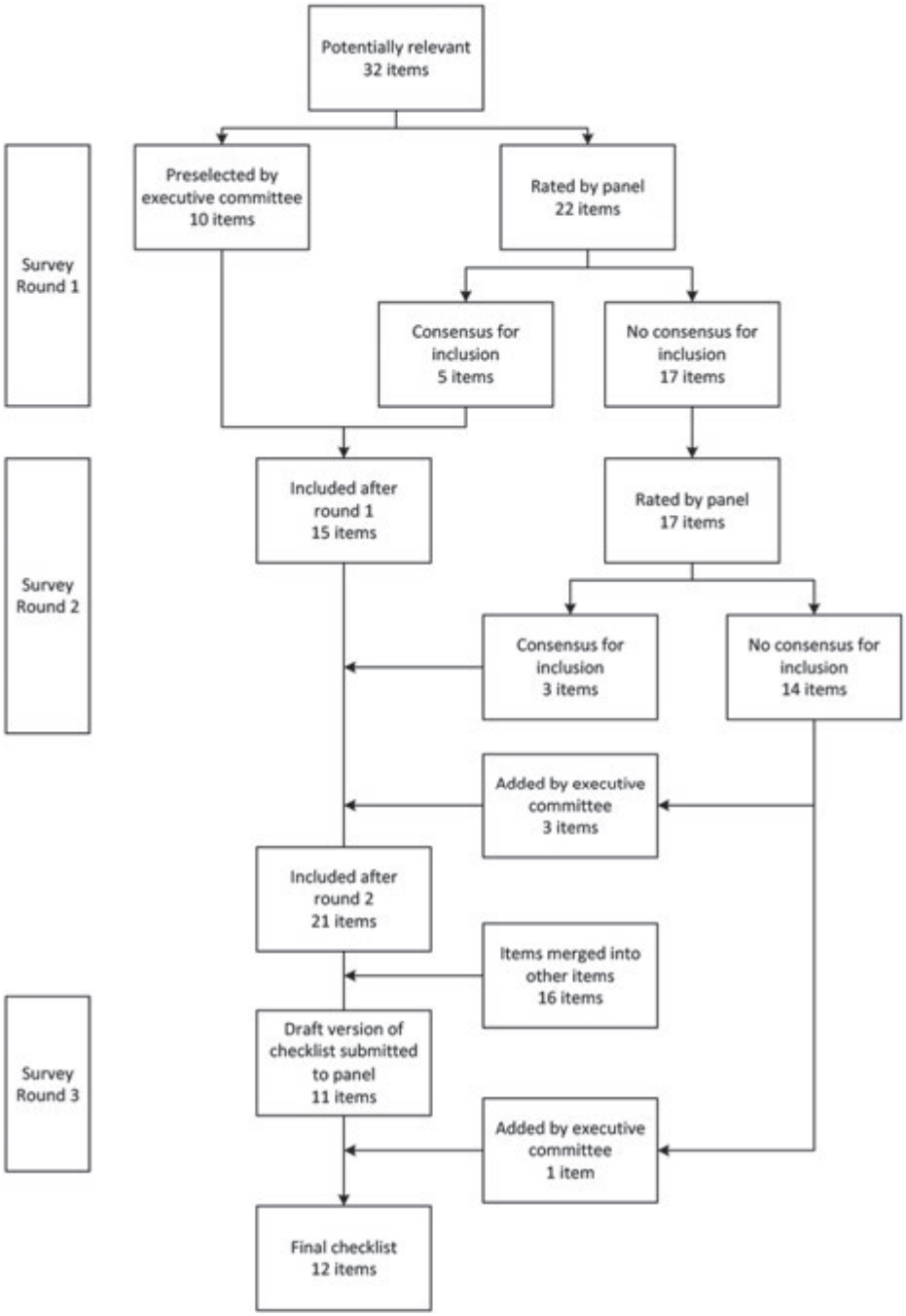
Slightly adapted from:
*Bertens LC, Reitsma JB, Moons KG, van Mourik Y, Lammers JW, Broekhuizen BD, Hoes AW, Rutten FH. Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis. 2013;8:493-9. doi: 10.2147/COPD.S49609*

**Supplemental Table 6. Comparison of the items of TRIPOD and TRIPOD for Abstracts**

| TRIPOD | TRIPOD for Abstracts |
|---|---|
| **Title and abstract** | ***Title*** |
| *1.* ***Title***: Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted. | 1. Identification of the study as developing, validating, or updating a prediction model, the target population, and the outcome to be predicted. |
| *1.* ***Abstract***: Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions. | |
| **Introduction** | |
| *2.* ***Background and objectives*** | ***Background*** |
| a. Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models. | 2. A brief explanation of the healthcare context (including whether diagnostic or prognostic) and rationale for developing, validating, or updating the model. |
| b. Specify the objectives, including whether the study describes the development or validation of the model or both. | ***Objectives*** |
| a. | 3. Study objectives, including whether the study describes the development, validation, or updating of a model. For validation of an existing model, give the name or describe the model being validated. |
| **Methods** | ***Methods*** |
| *3.* ***Source of data*** | 4. Study design or source of data (e.g., cohort, registry, routine care data, randomized trial), separately for the development and validation data sets, if applicable. |
| c. Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable. | |
| d. Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up | |
| *4.* ***Participants*** | 5. Participant eligibility criteria and setting where the data were collected. |
| a. Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres. | |
| b. Describe eligibility criteria for participants. | |
| c. Give details of treatments received, if relevant | |

| | |
|---|---|
| 5. ***Outcome***<br>a. Clearly define the outcome that is predicted by the prediction model, including how and when assessed.<br>b. Report any actions to blind assessment of the outcome to be predicted. | 6. Outcome to be predicted by the model, including time horizon of predictions in case of prognostic models (e.g., 3-year overall survival). |
| 6. ***Predictors***<br>a. Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.<br>b. Report any actions to blind assessment of predictors for the outcome and other predictors. | |
| 7. ***Sample size***: Explain how the study size was arrived at. | |
| 8. ***Missing data***: Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method. | |
| 9. ***Statistical analysis methods***<br>a. Describe how predictors were handled in the analyses.<br>b. Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.<br>c. For validation, describe how the predictions were calculated.<br>d. Specify all measures used to assess model performance and, if relevant, to compare multiple models.<br>e. Describe any model updating (e.g., recalibration) arising from the validation, if done. | 7. Statistical model or algorithm used (e.g. logistic regression, Cox regression, random forest, neural network) and approach for internal validation (for development studies). |
| 10. ***Risk groups***: Provide details on how risk groups were created, if done. | |
| 11. ***Development vs. validation***: For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors. | |
| **Results** | ***Results*** |
| 12. ***Participants***<br>a. Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.<br>b. Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome. | |

| | |
|---|---|
| c. For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome). | |
| *13.* ***Model development*** | 8. Number of participants and outcome events. |
| a. Specify the number of participants and outcome events in each analysis. | |
| b. If done, report the unadjusted association between each candidate predictor and outcome. | |
| *14.* ***Model specification*** | 9. Predictors in the final model (for development studies). |
| a. Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point). | |
| b. Explain how to use the prediction model. | |
| *15.* ***Model performance***: Report performance measures (with CIs) for the prediction model. | 10. Performance measures, at least calibration and discrimination (with confidence intervals), and results for added value of predictors or for model-updating, if applicable. |
| *16.* ***Model-updating***: If done, report the results from any model updating (i.e., model specification, model performance). | |
| **Discussion** | ***Discussion*** |
| *17.* ***Limitations***: Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data). | 11. Overall interpretation of the results, including implications for practice or research. |
| *18.* ***Interpretation*** | |
| a. For validation, discuss the results with reference to performance in the development data, and any other validation data. | |
| b. Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence. | |
| *19.* ***Implications***: Discuss the potential clinical use of the model and implications for future research. | |
| **Other information** | |
| *20.* ***Supplementary information***: Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets. | |
| *21.* ***Funding:*** Give the source of funding and the role of the funders for the present study. | |
| | ***Registration*** |
| | 12. Registration number and name of registry or repository. |




**Supplemental Figure 1. Item flow during development of TRIPOD for Abstracts**

**Supplement 3 – Examples of adequate reporting in abstracts**

*Example 1. Development and validation of a prognostic prediction model*

**Prediction of complications in early-onset pre-eclampsia (PREP): development and external multinational validation of prognostic models.**

**Background** Unexpected clinical deterioration before 34 weeks gestation is an undesired course in early-onset pre-eclampsia. To safely prolong preterm gestation, accurate and timely prediction of complications is required. We developed and externally validated multivariable prognostic models for providing individual risks of adverse maternal outcomes in women with early-onset pre-eclampsia, by 48 hours and by discharge.

**Method** Women with confirmed early onset pre-eclampsia were recruited from 53 maternity units in the UK to a prospective cohort study (PREP-946) for development of prognostic models for the overall risk of experiencing a complication using logistic regression (PREP-L), and for predicting the time to adverse maternal outcome using a survival model (PREP-S). For internal validation we used non-parametric bootstrapping to estimate over-optimism in performance. External validation of the models was carried out in a multinational cohort (PIERS-634, n=636) and another cohort from the Netherlands (PETRA-216, n=216).

**Results** In the PREP dataset 169 mothers (18%) had adverse outcomes by 48 hours, and 633 (67%) by discharge. The C-statistics of the models for predicting complications by 48 hours and by discharge were 0.84 (95% CI, 0.81-0.87; PREP-S) and 0.82 (0.80-0.84; PREP-L), respectively. The PREP-S model included maternal age, gestation, medical history, systolic blood pressure, deep tendon reflexes, urine protein creatinine ratio, platelets, serum alanine amino transaminase, urea, creatinine, oxygen saturation and treatment with antihypertensives or magnesium sulfate. The PREP-L model included the above except deep tendon reflexes, serum alanine amino transaminase and creatinine. On validation in the external PIERS dataset, the reduced PREP-S model showed reasonable calibration (slope 0.80) and discrimination (C-statistic 0.75, 95% CI, 0.69–0.81) for predicting adverse outcome by 48 hours. Reduced PREP-L model showed excellent calibration (slope: 0.93 PIERS, 0.90 PETRA) and discrimination (0.81 [0.77–0.85]) PIERS; 0.75 [0.64–0.86] PETRA) for predicting risk by discharge in the two external datasets.

**Conclusions** PREP models can be used to obtain predictions of adverse maternal outcome risk, including early preterm delivery, by 48 hours (PREP-S) and by discharge

(PREP-L), in women with early onset pre-eclampsia in the context of current care. They have a potential role in triaging high-risk mothers who may need transfer to tertiary units for intensive maternal and neonatal care.

**Trial registration** ISRCTN40384046.

Slightly adapted from: *Thangaratinam S, Allotey J, Marlin N et al. Prediction of complications in early-onset pre-eclampsia (PREP): development and external multinational validation of prognostic models. BMC Med 2017, 15(1): 68*
(word count n=362)

*Example 2. External validation of a prognostic prediction model*
**External validation of the updated ADO Score for predicting mortality in COPD patients from the Birmingham COPD Cohort.**

**Background** Reviews suggest that the ADO score is the most discriminatory prognostic score for predicting mortality among chronic obstructive pulmonary disease (COPD) patients, but a full evaluation and external validation within primary care settings is critical before implementation.

**Objectives** To validate the ADO score in prevalent and screen-detected primary care COPD cases at 3 years and at shorter time periods.

**Patients and methods** One thousand eight hundred and ninety-two COPD cases were recruited between 2012 and 2014 from 71 United Kingdom general practices as part of the Birmingham COPD Cohort study. Cases were either on the practice COPD register or screen-detected. We validated the ADO score for predicting 3-year mortality with 1-year and 2-year mortality as secondary endpoints using discrimination (area-under-the-curve (AUC)) and calibration plots.

**Results** One hundred and fifty-four deaths occurred within 3 years. The ADO score was discriminatory for predicting 3-year mortality (AUC= 0.74; 95% CI: 0.69-0.79). Similar performance was found for 1- (AUC= 0.73; 0.66-0.80) and 2-year mortality (0.72; 0.67-0.76). The ADO score showed reasonable calibration for predicting 3-year mortality (calibration slope 0.95; 0.70-1.19) but over-predicted in cases with higher predicted risks of mortality at 1 (0.79; 0.45-1.13) and 2-year (0.79; 0.57-1.01) mortality.

**Discussion** The ADO score showed promising discrimination in predicting 3-year mortality in a primary care population including screen-detected cases. It may need to be recalibrated if it is used to provide risk predictions for 1- or 2-year mortality since, in these time-periods, over-prediction was evident, especially in cases with higher predicted mortality risks.

Slightly adapted from: *Keene SJ, Jordan RE, Franssen FM, de Vries F, Martin J, Sitch A, Turner AM, Dickens AP, Fitzmaurice D, Adab P. External Validation Of The Updated ADO Score In COPD Patients From The Birmingham COPD Cohort. Int J Chron Obstruct Pulmon Dis. 2019 Oct 24;14:2395-2407. doi: 10.2147/COPD.S212381*
(word count: 249)

*Example 3. External validation and updating of a prognostic prediction model*
**Validation of a prediction model for long-term outcome of aphasia after stroke.**



**Background** About 30% of stroke patients suffer from aphasia. As aphasia strongly affects daily life, most patients request a prediction of outcome of their language function. Prognostic models provide predictions of outcome, but external validation is essential before models can be used in clinical practice. We aim to externally validate the prognostic model from the Sequential Prognostic Evaluation of Aphasia after stroKe (SPEAK-model) for predicting the long-term outcome of aphasia caused by stroke.

**Methods** We used data from the Rotterdam Aphasia Therapy Study - 3 (RATS-3), a multicenter RCT with inclusion criteria similar to SPEAK, an observational prospective study. Baseline assessment in SPEAK was four days after stroke and in RATS-3 eight days. Outcome of the SPEAK-model was the Aphasia Severity Rating Scale (ASRS) at 1 year, dichotomized into good (ASRS-score of 4 or 5) and poor outcome (ASRS-score < 4). In RATS-3, ASRS-scores at one year were not available, but we could use six month ASRS-scores as outcome. Model performance was assessed with calibration and discrimination.

**Results** We included 131 stroke patients with first-ever aphasia. At six months, 86 of 124 (68%) had a good outcome, whereas the model predicted 88%. Discrimination of the model was good with an area under the receiver operation characteristic curve of 0.87 (95%CI: 0.81-0.94), but calibration was unsatisfactory. The model overestimated the probability of good outcome (calibration-in-the-large $\alpha = -1.98$) and the effect of the predictors was weaker in the validation data than in the derivation data (calibration

slope β = 0.88). We therefore recalibrated the model to predict good outcome at six months.

**Conclusion** The original model, renamed SPEAK-12, has good discriminative properties, but needs further external validation. After additional external validation, the updated SPEAK-model, SPEAK-6, may be used in daily practice to discriminate between patients with good and patients with poor outcome of aphasia at six months after stroke.

**Trial registration** RATS-3 was registered on January 13th 2012 in the Netherlands Trial Register: NTR3271. SPEAK was not listed in a trial registry.

From: *Nouwens F, Visch-Brink EG, El Hachioui H, Lingsma HF, van de Sandt-Koenderman MWME, Dippel DWJ, Koudstaal PJ, de Lau LML. Validation of a prediction model for long-term outcome of aphasia after stroke. BMC Neurol. 2018 Oct 15;18(1):170. doi: 10.1186/s12883-018-1174-5.*
(word count: 326)

*Example 4. Development of a diagnostic prediction model*
**Development of a risk score for significant colonic pathology to stratify symptomatic adults referred for colonoscopy.**

**Background and aim** With an increasing burden on overstretched colonoscopy services, a simple risk score for significant pathology in symptomatic patients may aid in the prioritization of patients. We developed a diagnostic scoring system for significant colonic pathology in a multi-ethnic Asian population with symptoms.

**Methods** A cross-sectional study was conducted in consecutive symptomatic adults from an urban population referred for an index colonoscopy. Outcomes of interest were colonic neoplasia (colorectal carcinoma [CRC] and advanced adenoma) and CRC alone. The accuracy of the final model was assessed by the area under the curve (AUC) of the receiver operating characteristic curve and the Hosmer-Lemeshow goodness-of-fit statistic.

**Results** A total of 1013 subjects (mean age 59.9 ± 13.7 years, 52.3% females) from a multi-ethnic Asian background (Chinese 56%, Malay 20.4%, Indian 21.5%) were recruited. Colonic neoplasia and CRC were identified in 175 (17.3%) and 114 (11.3%) cases, respectively. Risk scores were assigned to individual factors identified in a logistic regression model of both demographic (age, gender, ethnicity, education level,

smoking history, Aspirin use) and clinical symptoms (change in bowel habit, bloody stool, weight loss, appetite loss, lethargy). The risk score for each patient was the sum of their individual risk factors. The AUC of the risk score for colonic neoplasia and CRC was 0.76 [0.72-0.80] (Hosmer-Lemeshow goodness-of-fit statistic of P = 0.745) and 0.83 [0.79-0.87] (Hosmer-Lemeshow goodness-of-fit statistic of P = 0.982), respectively.

**Conclusion** A simple risk score for colonic neoplasia and CRC may be able to prioritize colonoscopy referrals in symptomatic subjects from a multi-ethnic background. A further study to validate this scoring system is required.

Slightly adapted from: *Law CW, Rampal S, Roslani AC, Mahadeva S. Development of a risk score to stratify symptomatic adults referred for colonoscopy. J Gastroenterol Hepatol. 2014 Nov;29(11):1890-6. doi: 10.1111/jgh.12638.*
(word count: 259)



*Example 5. Development of a prognostic prediction model using machine learning*
**Training machine learning models to predict 30-day mortality in patients discharged from the emergency department: a retrospective, population based registry study**

**Objectives** Buying into the hypothesis that patients who are given an opportunity to communicate their end of life (EOL) preferences are more likely to receive EOL care in line with their preferences, the aim of this work was to train machine learning models to identify patients at EOL with clinically meaningful diagnostic accuracy, using 30-day mortality in patients discharged from the emergency department (ED) as a proxy.

**Design** Retrospective, population-based registry study.

**Setting** Swedish health services.

**Primary and secondary outcome measures** All cause 30-day mortality.

**Methods** Electronic health records (EHRs) and administrative data, including age, gender, comorbidities, whether referred by a physician, transported to ED in ambulance, urgency of medical condition, radiology order occurring during ED visit, and moment of discharge, were used to train six supervised machine learning models to predict all-cause mortality within 30 days in patients discharged from EDs in southern Sweden, Europe.

**Participants** The models were trained using 65 776 ED visits and validated on 55 164 visits from a separate ED to which the models were not exposed during training.

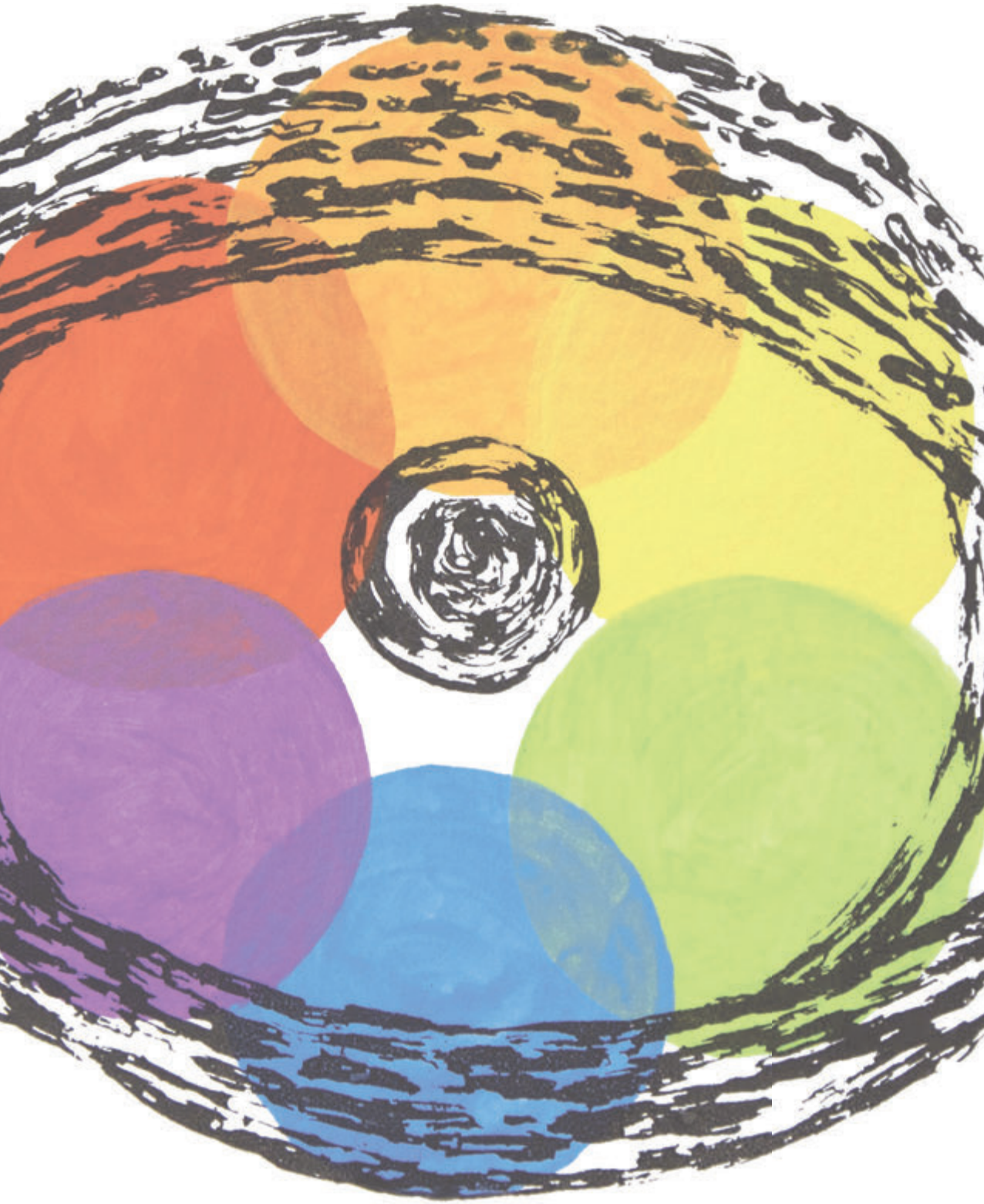
**Results** The outcome occurred in 136 visits (0.21%) in the development set and in 83 visits (0.15%) in the validation set. The model with highest discrimination attained ROC–AUC 0.95 (95% CI 0.93 to 0.96), with sensitivity 0.87 (95% CI 0.80 to 0.93) and specificity 0.86 (0.86 to 0.86) on the validation set.

**Conclusions** Multiple models displayed excellent discrimination on the validation set and outperformed available indexes for short-term mortality prediction in terms of ROC–AUC (by indirect comparison). The practical utility of the models increases as the data they were trained on did not require costly de novo collection but were real-world data generated as a by-product of routine care delivery.

Slightly adapted from: *Blom MC, Ashfaq A, Sant'Anna A, et al. Training machine learning models to predict 30-day mortality in patients discharged from the emergency department: a retrospective, population-based registry study. BMJ Open 2019;9:e028015. doi:10.1136/bmjopen-2018-028015*
(word count: 287)

**References**


1. Heus P, Damen J, Pajouheshnia R, Scholten R, Reitsma JB, et al. Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. *BMC Med*. 2018;16(1):120.
2. Heus P, Damen J, Pajouheshnia R, Scholten R, Reitsma JB, et al. Uniformity in measuring adherence to reporting guidelines: the example of TRIPOD for assessing completeness of reporting of prediction model studies. *BMJ Open*. 2019;9(4):e025611.
3. Nair R, Aggarwal R, Khanna D. Methods of formal consensus in classification/diagnostic criteria and guideline development. *Semin Arthritis Rheum*. 2011;41(2):95-105.
4. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess*. 1998;2(3):i-iv, 1-88.
5. SurveyMonkey Inc. www.surveymonkey.com. Accessed2020.
6. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63.
7. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-73.
8. Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, et al. PRISMA for Abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med*. 2013;10(4):e1001419.
9. Cohen JF, Korevaar DA, Gatsonis CA, Glasziou PP, Hooft L, et al. STARD for Abstracts: essential items for reporting diagnostic accuracy studies in journal or conference abstracts. *BMJ*. 2017;358:j3751.
10. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, et al. CONSORT for reporting randomised trials in journal and conference abstracts. *Lancet*. 2008;371(9609):281-283.



# Chapter 5

## How to promote the use of reporting guidelines: endorsement of TRIPOD and findings from an online survey among journal editors

Pauline Heus
Emma Kruithof
Johanna AAG Damen
Maud S Verhoef-Jurgens
Rob JPM Scholten
Johannes B Reitsma
Karel GM Moons
Lotty Hooft

*Manuscript in preparation*

## Abstract

### Background

To improve the value of biomedical research, numerous reporting guidelines have been developed. The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement is one of them. We aimed to assess endorsement of TRIPOD and of reporting guidelines in general by medical journals, and to identify journal editors' opinions and experiences regarding promoting the use of reporting guidelines.

### Methods

We selected the top 10 journals with the highest journal impact factor within each of 37 clinical domains and searched their online 'Instructions to authors' in February 2017 and in October 2018 for any reference to TRIPOD or other guidelines. We invited the editors-in-chief to participate in an online survey on the journal's editorial policies regarding reporting guidelines, and on (potential) barriers and facilitators to endorsement and active use of these guidelines.

### Results

In 2017, 205 out of 337 (61%) journals mentioned any reporting guideline in their instructions to authors. A reference to TRIPOD was provided by 27 (8%) journals. For 2018 these numbers were 219 (65%) and 29 (9%), respectively. Of those journals mentioning TRIPOD, 34% provided a link to the checklist. None of the journals required the use of TRIPOD.

Sixteen percent of journals (52/333) participated in our survey and 44% (18/41) was familiar with TRIPOD. Lack of knowledge among authors, reviewers, and editors; putting a burden on authors and peer reviewers; inflexibility; fear of less submissions; and the large number of available reporting guidelines, were identified as potential barriers to using guidelines.

### Conclusion

About two thirds of medical journals endorse reporting guidelines and 9% endorses TRIPOD. Journal editors suggested various actions to improve the use of reporting guidelines: journals requiring guideline use by authors; education and dissemination of tools, to all stakeholders and preferably centrally organized; and the use of automated tools to select the relevant guideline and check compliance.

## Introduction

Complete and accurate research reports enable clinicians, researchers, and other readers to make optimal use of the available evidence. Without a clear description of the research question addressed, the methods used, the results and implications, the usability of research is reduced and the research efforts can be considered as less valuable.[1,2]

To prevent this form of research waste and assist researchers in writing transparent and informative reports, reporting guidelines have been developed. A reporting guideline is defined as a checklist, flow diagram, or structured text to guide authors in reporting a specific type of research, developed using explicit methodology.[3] Many reporting guidelines exist for various types of study designs. The CONsolidated Standards Of Reporting Trials (CONSORT) statement, Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, Strengthening the Reporting of OBservational studies in Epidemiology (STROBE) statement, and STAndards for Reporting of Diagnostic Accuracy (STARD) statement are well-known examples.[4-7] A comprehensive collection of reporting guidelines is maintained by the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network, an international collaboration launched in 2008 that aims to promote responsible reporting of health research by providing resources and training, and by assisting in reporting guideline development, dissemination, and implementation.[8,9]



To promote the use of a reporting guideline (implementation) more is needed than just its publication.[10] One of the recommended post-publication activities is encouraging medical journals to support the use of the reporting guideline by incorporating it in their editorial policies and instructions to authors. Such explicit support (endorsement) was associated with more complete reporting for CONSORT, yet, for other reporting guidelines, to date the evidence is lacking.[11,12]

In 2015 the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement was published, a reporting guideline aiming to improve the completeness and transparency of diagnostic and prognostic prediction model reports.[13,14] We aimed to assess endorsement by medical journals of TRIPOD in particular and of reporting guidelines in general, and to identify journal editors' opinions and experiences regarding promoting the use of reporting guidelines.

## Methods

We selected the top 10 journals with the highest journal impact factor within each of the 37 clinical domains (subject categories, 2012 Journal Citation Reports ®)[15]

These journals were the starting point of our study which consisted of two parts: an assessment of the instructions to authors on the journals' websites, and a web-based survey among the journal editors.

**Assessment of 'Instructions to authors' on journals' websites**

For each journal we searched the instructions to authors for information on reporting guidelines in general and TRIPOD in particular. The following search terms were used: *'reporting', 'guideline', 'statement', 'checklist', 'endorse', 'EQUATOR',* '*TRIPOD',* and *'CONSORT'*. Since there are over 400 reporting guidelines, it was not possible to search for every guideline separately.[3] Although we extracted information on any mentioned reporting guideline, CONSORT was explicitly included in the search terms because it is one of the oldest reporting guidelines and is highly cited and endorsed.[16,17] Links in the instructions for authors to other locations on the journal's website or to other websites were followed if they seemed relevant to reporting and information presented there was included. In the case of different journals providing the same instructions, these were included for every individual journal separately.

We extracted information on which reporting guidelines were mentioned and whether the EQUATOR Network was acknowledged. We also checked whether the journal provided a functioning link to additional information regarding these reporting guidelines or the EQUATOR Network. Furthermore, with regard to TRIPOD, we noted which source of additional information was referenced (website, publication, checklist, other) and whether adhering to TRIPOD was required (using explicit language, like 'authors must follow', 'authors are required to'); recommended (using less insistent wording, like 'authors should adhere to', 'authors are recommended to use'); or suggested (providing authors the option by statements like 'authors can follow', 'authors are encouraged to use').

One author (PH, JAAGD, EK, or MSV-J) assessed the instructions to authors on the journals' websites between November 28th, 2016, and February 26th, 2017, and again between July 25th and October 31st, 2018. A second author checked the websites of the journals for which information regarding reporting guidelines was not identified (anymore).

Data were summarized descriptively using frequencies and percentages.

**Survey among journal editors**

To elicit journal policies and journal editors' opinions and experiences regarding endorsement and implementation of TRIPOD and other reporting guidelines, we used an online software tool to develop and run a web-based survey.[18] A schematic

representation of the survey is provided in Figure 1. The survey included both multiple choice questions and open-ended questions and was strictly anonymous. For administrative purposes we asked respondents to provide the name of their journal, however, this was optional. Before inviting editors to participate, the survey was piloted by the author team.

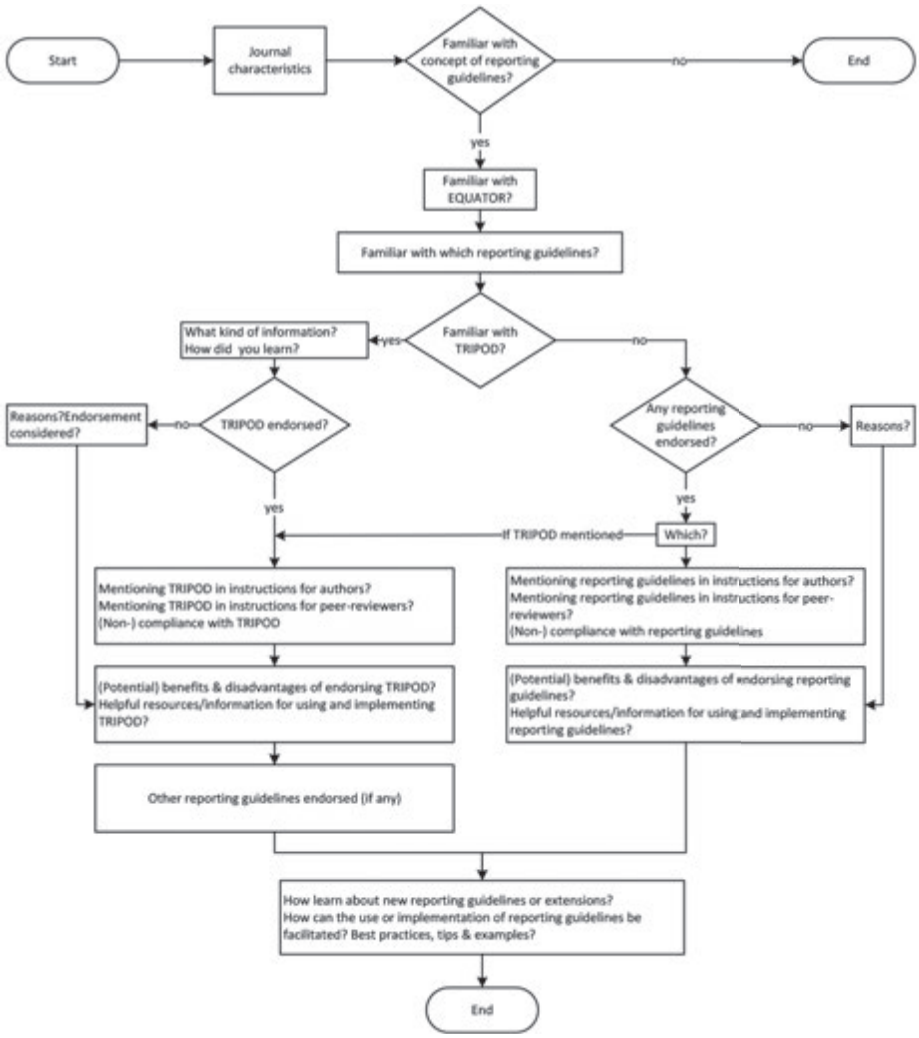



**Figure 1. Schematic flow of questions within the web-based survey amongst journal editors**

We invited the editor-in-chief of each selected journal to participate in the survey. Invitations were sent by e-mail containing a web link to the survey. Contact details of

the editor-in-chief or editorial office were obtained from either the journal's website or a name-based internet search. In case we retrieved multiple e-mail addresses, we sent the invitation to all of these in order to increase the likelihood of a response. The initial invitation was followed by reminders after two weeks and after another week. Editors-in-chief received an invitation on the 27th of September, 2018. The survey was open for response up to the 30th of October, 2018.

Available information from incomplete surveys was also included in the analysis. We used frequencies and percentages to summarize the data. Two authors (EK, PH) qualitatively analysed the answers to the open-ended questions, most of these were follow-up inquiries to multiple-choice questions.

## Results

Of the 370 journals selected, 341 unique journals remained after deduplication (Figure 2, Supplemental Table 1). Four journals were excluded because we were unable to identify a journal website with up-to-date information. This resulted in a set of 337 included journals with a median impact factor of 4.5 (25th–75th percentile [$P_{25}$–$P_{75}$]: 3.2-7.1).

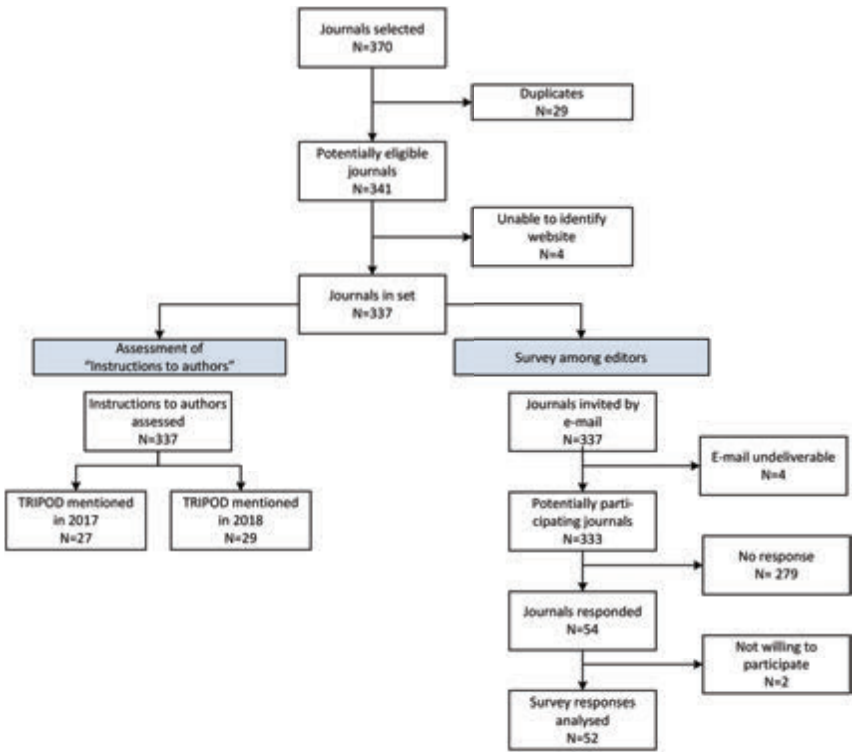



**Figure 2. Flow of journals through the study**

**Assessment of 'Instructions to authors'**

The number of journals mentioning any reporting guidelines in their instructions to authors increased slightly from 205 (61%) in 2017 to 219 (65%) in 2018. Also the EQUATOR Network was mentioned by more journals in 2018 (102; 30%) compared to 2017 (79; 23%). The reporting guideline most frequently listed by the journals, in 2017 as well as in 2018, was CONSORT (2018: 178; 53%), followed by PRISMA (2018: 141; 42%), and STROBE (2018: 107; 32%) (Supplemental Table 2). Of the 226 journals mentioning any reporting guideline or the EQUATOR network in 2018, 216 (96%) provided a functioning web link to additional information compared to 175 of the 206 journals (85%) in 2017.

TRIPOD was mentioned by 27 (8%) journals in 2017 and 29 (9%) journals in 2018. Twenty-one journals mentioned TRIPOD in both years, so six journals mentioning TRIPOD in 2017 did not do so anymore in 2018. Journal impact factor and the clinical domains in which the journals published were similar for both sets (Supplemental Table 3).



**Table 1. Details regarding TRIPOD resources referenced and guidance provided by the journals mentioning TRIPOD in their instructions to authors**

| | 2017 (n=27 journals) | 2018 (n=29 journals) |
|---|---|---|
| **Rescources referenced*** | | |
| TRIPOD website | 2 (7%) | 4 (14%) |
| TRIPOD checklist | 8 (30%) | 10 (34%) |
| TRIPOD statement paper | 2 (7%) | 3 (10%) |
| TRIPOD explanatory paper | 0 (0%) | 0 (0%) |
| TRIPOD information on EQUATOR Network website | 7 (26%) | 5 (17%) |
| EQUATOR Network website homepage | 19 (70%) | 22 (78%) |
| **Guidance*** | | |
| Obligation to follow TRIPOD or provide completed checklist | 0 (0%) | 1 (3%) |
| Recommendation to follow TRIPOD or provide completed checklist | 12 (44%) | 12 (41%) |
| Suggestion to follow TRIPOD or provide completed checklist | 9 (33%) | 10 (34%) |
| General recommendation to consult EQUATOR Network | 21 (78%) | 21 (72%) |
| No TRIPOD specific guidance, nor referral to EQUATOR Network | 4 (15%) | 4 (14%) |

Number of journals (%); EQUATOR: Enhancing the QUAlity and Transparency Of health Research; TRIPOD: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

*Numbers add up to over 29, as more than one category could apply to a journal.

Ten (34%) of the 29 journals mentioning TRIPOD in 2018 provided a web link to the TRIPOD checklist (Table 1). Five journals (17%) had a link to TRIPOD information on the website of the EQUATOR Network and four had a link to the TRIPOD website (14%). Three journals (10%) referenced the publication of the TRIPOD statement. A reference to the general homepage of the EQUATOR website was provided by 22 journals (78%).

With regard to the type of guidance provided by the journals, there was one journal (3%) that required authors to upload a completed TRIPOD checklist. All other journals used less explicit language and recommended (12 journals; 41%) or suggested (10 journals; 34%) to follow TRIPOD or complete its checklist. A general recommendation to consult the EQUATOR Network was given by 21 journals (72%). Four journals (14%) did not provide any guidance regarding TRIPOD or EQUATOR. The results for the journals mentioning TRIPOD in 2017 were comparable to those in 2018 (Table 1).

Eight of the eleven journals that published the TRIPOD statement in 2015 were included in our set. All but one of these (88%) mentioned reporting guidelines or EQUATOR in their instructions to authors and provided web links. Only three of them (38%) mentioned TRIPOD in their instructions to authors .

**Survey among journal editors**

Of the 337 invitations, four proved to be undeliverable, two journals did not want to participate and 279 did not reply (Figure 2), leaving 52 survey responses (52/333; 16%). Seven of these responses (13%) were incomplete.

*Journal and respondent characteristics*

Most responding journals were specialized journals (39; 75%, Supplemental table 4). Their median journal impact factor was 4.3 ($P_{25}$–$P_{75}$: 2.8-6.9). Forty nine respondents provided information on the study types their journal publishes: most mentioned were systematic reviews (92%), followed by observational studies (82%), and randomised trials (76%). The majority of the respondents were editors-in-chief (71%), and most of them were familiar with reporting guidelines (88%) and the EQUATOR Network (81%). CONSORT (90%), PRISMA (88%), STROBE (88%), and STARD (51%) were the most well-known reporting guidelines.

The EQUATOR Network (58%) and colleagues (56%) were most often mentioned as ways to learn about new reporting guidelines or extensions to existing guidelines.

Eighteen editors (of 41 respondents; 44%) were aware of TRIPOD, of which 16 (89%) were familiar with the checklist and 13 (72%) with the TRIPOD statement. The TRIPOD website (22%) and explanation and elaboration paper (17%) were less well known. In most cases they learned about TRIPOD through colleagues (44%), followed by the EQUATOR Network (39%), authors (22%), the TRIPOD statement (6%) or conferences (6%). Of all 36 journals publishing diagnosis and prognosis research, 16 editors (44%) indicated that they were not familiar with TRIPOD.

*Factors related to endorsement of reporting guidelines and editorial policies*
Of the 41 respondents, 35 (85%) endorsed reporting guidelines, with CONSORT (88%), STROBE (76%), and PRISMA (74%) as the top 3 of most endorsed guidelines (Supplemental table 4). TRIPOD was endorsed by 24% of the journals. As reasons for not endorsing TRIPOD, respondents stated that the process of adopting reporting guidelines takes time and indicated that they were currently planning to endorse TRIPOD. Lack of experience with TRIPOD was another reason. Publishing study types for which reporting guidelines are not applicable and lack of knowledge were also pointed out with regard to endorsement of reporting guidelines in general. Some journals explicitly indicated not to endorse guidelines and leave it to the authors and peer reviewers. One respondent acknowledged the (to him or her unfounded) fear of the editorial leadership that adhering to reporting guidelines would depress submissions.

A summary of the survey responses regarding editorial policies is provided in Table 2. Journals most often refer authors to the website of the reporting guideline (60%), directly to the checklist (25%), or more generic to EQUATOR (25%). To peer reviewers, in 36% no specific tools were offered. Editorial teams checked mainly through authors submitting a checklist (47%) or providing a statement (37%) whether a publication complied with a reporting guideline.

**Table 2. Summary of survey responses on editorial policies regarding reporting guidelines**

| | N | n (%) |
|---|---|---|
| **TRIPOD / reporting guidelines are mentioned in the instructions to authors.** | **27** | **20 (74%)** |
| **What tools do you offer to authors?*** | **20** | |
| No tools are offered to authors | | 1 (5%) |
| Link to reporting guideline's website | | 12 (60%) |
| Link to EQUATOR website | | 5 (25%) |
| Explanatory document | | 3 (15%) |
| Online tutorial | | 0 (0%) |
| Checklist | | 5 (25%) |
| Automated screening of manuscripts | | 1 (5%) |
| **TRIPOD / reporting guidelines are mentioned in the instructions to peer reviewers.** | **26** | **11 (42%)** |
| **What tools do you offer to peer reviewers?*** | **11** | |
| No tools are offered to peer reviewers | | 4 (36%) |
| Link to reporting guideline(s) website | | 3 (27%) |
| Link to EQUATOR website | | 0 (0%) |
| Explanatory document | | 0 (0%) |
| Online tutorial | | 0 (0%) |
| Checklist | | 2 (18%) |
| Automated screening of manuscripts | | 0 (0%) |
| **How does the editorial team check whether manuscripts comply with a reporting guideline?*** | **30** | |
| A statement of the authors is requested | | 11 (37%) |
| A checklist must be submitted | | 14 (47%) |
| Editors are asked to check | | 3 (10%) |
| Reviewers are asked to check | | 1 (3%) |
| It is not checked | | 3 (10%) |
| Other | | 2 (7%) |
| **What is the editorial policy for manuscripts suitable for publication but not compliant with the reporting guideline?*** | **30** | |
| They are accepted | | 3 (10%) |
| They are returned for revision | | 20 (67%) |
| It is not checked if manuscripts comply with reporting guideline(s) | | 5 (17%) |

N=number of respondents
*respondents could provide multiple answers
**Respondents explained that it is left to the reviewers and (associate) editors, but not actively asked or consistently reminded.

*Factors related to promoting the use of reporting guidelines*

Facilitators

Almost all respondents (95%) were convinced that reporting guidelines result in more complete reporting and in better quality of manuscripts (83%) (Table 3). The majority (75%) believed there is more need for dissemination and endorsement of reporting guidelines, because of these positive effects. In addition, they recognize the necessity to raise knowledge and awareness on the topic. Checklists (67%), an example study with complete and accurate reporting (57%), and online tutorials (55%) were mentioned as being most helpful to enhance the use of reporting guidelines. Additional suggestions included integration with automatic systems to select the relevant reporting guideline at submission or to label publications adhering to a specific reporting guideline; centralised resources; and educating publishers' employees.

**Table 3. Summary of survey responses on issues regarding implementation of reporting guidelines**

| | N | n (%) |
|---|---|---|
| **What are (potential) benefits of endorsing TRIPOD / reporting guidelines?*** | **42** | |
| More complete reporting in manuscripts | | 40 (95%) |
| Better quality of manuscripts | | 35 (83%) |
| Easier peer-review process | | 19 (45%) |
| Easier editorial process | | 21 (50%) |
| Other | | 2 (5%)** |
| **What are (potential) disadvantages of endorsing TRIPOD / reporting guidelines?*** | **42** | |
| It takes authors more time | | 18 (43%) |
| It takes reviewers more time | | 10 (24%) |
| Authors might prefer to publish in another journal not endorsing reporting guideline(s) | | 17 (40%) |
| A journal may have its own guidelines to adhere to | | 10 (24%) |
| Other | | 12 (29%)* |
| **There is more need for dissemination and endorsement of TRIPOD / reporting guidelines.**** | **40** | **30 (75%)** |
| **What type of resources or information would enhance the use and implementation of TRIPOD / reporting guidelines?*** | **42** | |
| Checklist | | 28 (67%) |
| Explanation & Elaboration document | | 17 (40%) |
| Website | | 18 (43%) |
| Template | | 21 (50%) |

| | |
|---|---|
| A sample study with examples | 24 (57%) |
| Application for electronic devices | 7 (17%) |
| Conference presentations | 15 (36%) |
| Online tutorials | 23 (55%) |
| Other | 5 (12%)** |

N=number of respondents
*respondents could provide multiple answers
**See article text for a summary of the explanations and comments provided

Barriers

Increased time needed for authors to prepare their manuscript was seen as disadvantage by 43% of the respondents and 40% thought that authors might prefer to publish in a journal not endorsing guidelines (Table 3). One of the themes emerging from the additional comments was that reporting guidelines lack flexibility and do not always fit well. In addition, not all authors, reviewers, and editors fully embrace endorsement, leading to disagreement on which reporting guideline to use or recommend and how the instructions should be formulated (e.g. encourage vs. require). Although endorsing reporting guidelines can be seen as a burden, several respondents stated that they did not see insuperable disadvantages. Some respondents were unsure whether there is need for more encouragement of reporting guidelines and they called for fewer reporting guidelines.

Sharing best practices

Respondents emphasized the importance of requiring and checking compliance to reporting guidelines by journals. Furthermore, education of various audiences (like PhD-students, editorial boards, and conference participants) was considered valuable. One respondent had good experiences with peer pressure after presenting general overviews of best reporting practices at a field's scientific meeting. Several ways to disseminate information were suggested, including editorials, instructions to authors, good websites, and article templates. Respondents emphasized the need for international consensus on core guidelines, because they feel there are currently too many guidelines. They also proposed a revision of the website of the EQUATOR Network, in order to make it easier to locate guidelines and download usable templates. Again, the potential of a submission platform that enables authors to automatically find the right checklist was mentioned.

## Discussion

About two thirds of medical journals endorse reporting guidelines by mentioning them in the journal's online instructions to authors. We noticed a slight increase from

61% in in February 2017 to 65% in October 2018. Most well-known and endorsed guidelines were CONSORT, PRISMA and STROBE (in 2018 mentioned by 53%, 42%, and 32% of the journals, respectively). In 2018, 9% of the journals mentioned TRIPOD. Most journals provided a link to the TRIPOD checklist, however, its use was recommended rather than required. Almost half of the editors participating in our survey were familiar with TRIPOD, mainly with the checklist. Potential barriers to endorsing reporting guidelines are lack of knowledge among authors, reviewers, and editors; the longer time authors and peer reviewers need when using a reporting guideline; inflexibility; fear of less submissions, as authors might prefer to submit to a non-endorsing journal; and the large number of reporting guidelines that currently exist.

Compared to other reporting guidelines, the percentage journals mentioning TRIPOD (9%) is low. However, TRIPOD is a relatively young reporting guideline that was published in 2015 and it is known that changing practice takes time. The first evaluation of endorsement of CONSORT by medical journals was performed seven years after its publication and showed that about 20% of high impact journals referred to it.[19] Moreover, CONSORT addresses randomised trials, a study design with a longer history than prediction model studies.



The evaluation of CONSORT endorsement has been repeated in 2007 and 2014 and showed an increase to 63% (in 2014) of high impact journals mentioning CONSORT in their author instructions.[17,20] CONSORT was mentioned less often by the journals in our set (53% in 2018). Although similarly broad, there were differences in the initial journal selection procedure between both studies (the CONSORT evaluation uses the top five impact factor journals for each of 33 medical specialties and the top 15 impact factor journals in general and internal medicine). Furthermore, the CONSORT evaluation excluded journals not likely to publish randomized trials.

Other assessments of instructions to authors in diverse clinical fields showed varying endorsement rates of mentioning CONSORT and other reporting guidelines.[21-27]. They did, however, agree on ambiguity in the guidance provided to authors, as journals were vague about to what extent adherence to reporting guidelines was required. For TRIPOD we found that only one journal required adherence. All other journals used less stringent wording and recommended or suggested to follow the TRIPOD guideline or checklist. In comparison, Shamseer et al. reported that in 2014 the use of CONSORT was required in 42% of high impact medical journals and that 53% recommended its use.[17]

There are several examples that a more active editorial strategy to implement reporting guidelines led to better adherence to reporting guidelines.[28,29] It is therefore interesting to find out what factors influence the policy of journals regarding reporting guidelines. Several studies surveying editors on this topic have been carried out.[30-32] Factors preventing endorsement found in these studies overlap with our results: lack of knowledge, putting a burden on authors and peer reviewers and the fear that authors will submit their manuscript to another journal.

A limitation of our study is that we used the complete set of journals for assessing the endorsement of TRIPOD, including journals that do not or hardly publish prediction model studies and thus have no reason to endorse TRIPOD. Therefore, a likely underestimation of endorsement of TRIPOD should be kept in mind when interpreting our results. A challenge regarding the assessment of the online instructions to authors was that journals changed their websites during the study period. In some cases, in 2018 even after double checking we could not find the information extracted in 2017 (e.g. six of the 27 journals mentioning TRIPOD in 2017 did not mention TRIPOD in 2018). Another limitation is the low response rate to our survey. The survey results reflect the view of a selected group of editors of journals with relatively high rates of endorsing reporting guidelines.

Despite these limitations, our results provide useful insights into potential ways to advance the implementation of TRIPOD and other reporting guidelines. Implementation would benefit from clear instructions to authors, as endorsement of reporting guidelines by medical journals is currently operationalized in various ways. Requiring adherence to reporting guidelines and checking author compliance are expected to enhance complete reporting (based on evidence, [33] as well as suggested by survey respondents). Nevertheless, regardless whether a journal requires adherence to reporting guidelines, authors have their own responsibility with regard to complete and transparent reporting of research findings and can use guidelines at any time.

The abundance of available reporting guidelines is a potential barrier to using them. The database of the EQUATOR Network currently holds 421 guidelines, and survey respondents believed it is not always easy to identify the most applicable guideline (extension) and tools for a particular study. It is possible that several reporting guidelines are applicable to a specific study, for example in the case of a randomized trial of a complex implementation intervention. In this situation CONSORT would apply, and also the Template for Intervention Description and Replication (TIDieR), the

Standards for Reporting Implementation Studies (StaRI) Statement, and the Standards for Quality Improvement Reporting Excellence (SQUIRE 2.0).[34-36]

A recent scoping review identified 31 interventions to improve adherence to reporting guidelines.[37] It is likely that software solutions will increasingly become available that can assist authors, peer reviewers, and editors in selecting the relevant guideline (e.g. the EQUATOR Wizard) and checking compliance with it (e.g. StatReviewer).[38,39] In addition, automatization will reduce the workload, which at present is another important barrier to using reporting guidelines.

Prerequisites for the use of reporting guidelines are awareness of their existence and access to available tools. This concerns not only authors, but also peer reviewers and editorial staff. Developers should keep all the various stakeholders in mind when disseminating their reporting guideline and developing educational materials and tools. In addition, the EQUATOR Network has an important, central role in providing resources and in making the selection of the pertinent reporting guideline more easy.



Raising awareness and providing education are especially important for TRIPOD, as it is a recent reporting guideline addressing a relatively young research field. According to the editors participating in our survey, there is need for good examples as useful educational tool, for TRIPOD more than for reporting guidelines in general (mentioned by 72% vs. 46% of the editors, respectively).

As the current study mainly represents the view of editors, future studies should explore authors' and peer reviewers' perspectives.

## Conclusion

About two thirds of medical journals endorse reporting guidelines, which is encouraging, as endorsement by journals is an important step in the implementation of reporting guidelines. Currently, 9% of the journals endorsed TRIPOD in their instructions to authors. Journal editors suggested various actions to improve the use of reporting guidelines, notably: journals requiring rather than recommending guideline use by authors; education and dissemination of tools on how to use reporting guidelines, to all stakeholders and preferably centrally organized; and the use of automated tools to assist in selecting the relevant guideline and checking compliance. Enhanced use of TRIPOD will promote adequate reporting of prediction model studies, making them more usable and thereby prevent research waste.

## References


1. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374(9683):86-89.
2. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267-276.
3. The EQUATOR Network | Enhancing the QUAlity and Transparency Of health Research. https://www.equator-network.org/. Accessed 2 February, 2020.
4. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152(11):726-732.
5. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
6. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573-577.
7. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
8. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. *Open Med*. 2008;2(2):e49-e50.
9. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med*. 2010;8:24.
10. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7(2):e1000217.
11. Stevens A, Shamseer L, Weinstein E, Yazdi F, Turner L, et al. Relation of completeness of reporting of health research to journals' endorsement of reporting guidelines: systematic review. *BMJ*. 2014;348.
12. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev*. 2012;11:Mr000030.
13. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63.
14. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-73.

15. 2012 Journal Citation Reports® In. Science Edition ed: Clarivate Analytics 2017.
16. Altman DG, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. *J R Soc Med.* 2016;109(2):67-77.
17. Shamseer L, Hopewell S, Altman DG, Moher D, Schulz KF. Update on the endorsement of CONSORT by high impact factor journals: a survey of journal "Instructions to Authors" in 2014. *Trials.* 2016;17(1):301.
18. SurveyMonkey Inc. www.surveymonkey.com. Accessed 2020.
19. Altman DG. Endorsement of the CONSORT statement by high impact medical journals: survey of instructions for authors. *BMJ.* 2005;330(7499):1056-1057.
20. Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT Statement by high impact factor medical journals: a survey of journal editors and journal 'Instructions to Authors'. *Trials.* 2008;9:20.
21. Smidt N, Overbeke J, de Vet H, Bossuyt P. Endorsement of the STARD Statement by biomedical journals: survey of instructions for authors. *Clin Chem.* 2007;53(11):1983-1985.
22. Tao KM, Li XQ, Zhou QH, Moher D, Ling CQ, Yu WF. From QUOROM to PRISMA: a survey of high-impact medical journals' instructions to authors and a review of systematic reviews in anesthesia literature. *PLoS One.* 2011;6(11):e27611.
23. Sharp MK, Tokalic R, Gomez G, Wager E, Altman DG, Hren D. A cross-sectional bibliometric study showed suboptimal journal endorsement rates of STROBE and its extensions. *J Clin Epidemiol.* 2019;107:42-50.
24. Meerpohl JJ, Wolff RF, Antes G, von Elm E. Are pediatric Open Access journals promoting good publication practice? An analysis of author instructions. *BMC Pediatr.* 2011;11:27.
25. Sims MT, Henning NM, Wayant CC, Vassar M. Do emergency medicine journals promote trial registration and adherence to reporting guidelines? A survey of "Instructions for Authors". *Scand J Trauma Resusc Emerg Med.* 2016;24(1):137.
26. Shantikumar S, Wigley J, Hameed W, Handa A. A survey of instructions to authors in surgical journals on reporting by CONSORT and PRISMA. *Ann R Coll Surg Engl.* 2012;94(7):468-471.
27. Smith TA, Kulatilake P, Brown LJ, Wigley J, Hameed W, Shantikumar S. Do surgery journals insist on reporting by CONSORT and PRISMA? A follow-up survey of 'instructions to authors'. *Ann Med Surg (Lond).* 2015;4(1):17-21.
28. Hopewell S, Ravaud P, Baron G, Boutron I. Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: interrupted time series analysis. *BMJ.* 2012;344:e4178.
29. Pandis N, Shamseer L, Kokich VG, Fleming PS, Moher D. Active implementation strategy of CONSORT adherence by a dental specialty journal improved randomized clinical trial reporting. *J Clin Epidemiol.* 2014;67(9):1044-1048.

30. Grindlay DJ, Dean RS, Christopher MM, Brennan ML. A survey of the awareness, knowledge, policies and views of veterinary journal Editors-in-Chief on reporting guidelines for publication of research. *BMC Vet Res.* 2014;10:10.
31. Fuller T, Pearson M, Peters J, Anderson R. What affects authors' and editors' use of reporting guidelines? Findings from an online survey and qualitative interviews. *PLoS One.* 2015;10(4):e0121585.
32. Hua F, Walsh T, Glenny AM, Worthington H. Surveys on Reporting Guideline Usage in Dental Journals. *J Dent Res.* 2016;95(11):1207-1213.
33. Agha RA, Fowler AJ, Limb C, Whitehurst K, Coe R, et al. Impact of the mandatory implementation of reporting guidelines on reporting quality in a surgical journal: A before and after study. *Int J Surg.* 2016;30:169-172.
34. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ.* 2014;348:g1687.
35. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, et al. Standards for Reporting Implementation Studies (StaRI) Statement. *BMJ.* 2017;356:i6795.
36. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf.* 2016;25(12):986-992.
37. Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. *BMJ Open.* 2019;9(5):e026589.
38. EQUATOR Network, Penelope Research. EQUATOR Wizard. https://www.goodreports.org/. Accessed 19th of February, 2020.
39. StatReviewer. http://www.statreviewer.com/. Accessed 19th of February, 2020.

## Supplementary material

**Supplemental Table 1. Selected journals (n=341)**

| Full Journal Title | Clinical domain* | Journal impact factor |
|---|---|---|
| **Academic Emergency Medicine** | Emergency medicine | 1.757 |
| **Acta Dermato-Venereologica** | Dermatology | 3.487 |
| **Acta Neuropathologica** | Clinical neurology | 9.734 |
| **Acta Orthopaedica** | Orthopedics | 2.736 |
| **Acta Tropica** | Tropical medicine | 2.787 |
| **Advances In Clinical Chemistry**** | Medical laboratory technology | 3.674 |
| **Age** | Geriatrics & Gerontology | 4.084 |
| **Ageing Research Reviews** | Geriatrics & Gerontology | 5.953 |
| **Aging Cell** | Geriatrics & Gerontology | 5.705 |
| **Aids** | Infectious diseases | 6.407 |
| **Allergy** | Allergy | 5.883 |
| **Allergy Asthma & Immunology Research** | Allergy | 2.653 |
| **Alternative Medicine Review**** | Integrative & complementary medicine | 4.857 |
| **Alzheimers & Dementia** | Clinical neurology | 14.483 |
| **American Family Physician** | Primary health care | 1.611 |
| **American Journal of Chinese Medicine** | Integrative & complementary medicine | 2.281 |
| **American Journal of Emergency Medicine** | Emergency medicine | 1.704 |
| **American Journal of Epidemiology** | Public, Environmental and Occupational health | 4.780 |
| **American Journal of Gastroenterology** | Gastroenterology & Hepatology | 7.553 |
| **American Journal of Geriatric Psychiatry** | Geriatrics & Gerontology | 4.131 |
| **American Journal of Kidney Diseases** | Urology & Nephrology | 5.294 |
| **American Journal of Obstetrics and Gynecology** | Obstetrics & Gynecology | 3.877 |
| **American Journal of Ophthalmology** | Ophthalmology | 3.631 |
| **American Journal of Physiology-Lung Cellular and Molecular Physiology** | Respiratory system | 3.523 |
| **American Journal of Psychiatry** | Psychiatry | 14.721 |
| **American Journal of Reproductive Immunology** | Reproductive biology | 3.317 |
| **American Journal of Respiratory and Critical Care Medicine** | Critical care medicine; Respiratory system | 11.041 |
| **American Journal of Respiratory Cell and Molecular Biology** | Respiratory system | 4.148 |
| **American Journal of Speech-Language Pathology** | Rehabilitation | 2.448 |
| **American Journal of Sports Medicine** | Orthopedics; Sport sciences | 4.439 |

| | | |
|---|---|---|
| **American Journal of Surgical Pathology** | Surgery | 4.868 |
| **American Journal of Transplantation** | Transplantation; Surgery | 6.192 |
| **American Journal of Tropical Medicine and Hygiene** | Tropical medicine | 2.534 |
| **Anaesthesia** | Anesthesiology | 3.486 |
| **Anesthesia and Analgesia** | Anesthesiology | 3.300 |
| **Anesthesiology** | Anesthesiology | 5.163 |
| **Annals of Allergy Asthma & Immunology** | Allergy | 3.449 |
| **Annals of Emergency Medicine** | Emergency medicine | 4.285 |
| **Annals of Family Medicine** | Primary health care | 4.613 |
| **Annals of Internal Medicine** | Medicine, general & internal | 13.976 |
| **Annals of Neurology** | Clinical neurology | 11.193 |
| **Annals of Surgery** | Surgery | 6.329 |
| **Annals of Surgical Oncology** | Surgery | 4.120 |
| **Annals of the Rheumatic Diseases** | Rheumatology | 9.111 |
| **Annals of Tropical Medicine and Parasitology** | Tropical medicine | 1.313 |
| **Annual Review of Immunology** | Immunology | 36.556 |
| **Antioxidants & Redox Signaling** | Endocrinology & Metabolism | 7.189 |
| **Archives of Dermatology** | Dermatology | 4.792 |
| **Archives of Disease In Childhood** | Pediatrics | 3.051 |
| **Archives of Disease In Childhood-Fetal and Neonatal Edition** | Pediatrics | 3.451 |
| **Archives of General Psychiatry** | Psychiatry | 13.772 |
| **Archives of Internal Medicine** | Medicine, general & internal | 10.579 |
| **Archives of Neurology** | Clinical neurology | 7.685 |
| **Archives of Ophthalmology** | Ophthalmology | 3.826 |
| **Archives of Otolaryngology-Head & Neck Surgery** | Otorhinolaryngology | 1.779 |
| **Archives of Pathology & Laboratory Medicine** | Medical laboratory technology | 2.781 |
| **Archives of Pediatrics & Adolescent Medicine** | Pediatrics | 4.282 |
| **Archives of Physical Medicine and Rehabilitation** | Rehabilitation | 2.358 |
| **Archives of Surgery** | Surgery | 4.100 |
| **Arteriosclerosis Thrombosis and Vascular Biology** | Hematology; Peripheral vascular disease | 6.338 |
| **Arthritis and Rheumatism** | Rheumatology | 7.477 |
| **Arthritis Care & Research** | Rheumatology | 3.731 |
| **Arthritis Research & therapy** | Rheumatology | 4.302 |
| **Arthroscopy-the Journal of Arthroscopic and Related Surgery** | Orthopedics | 3.103 |
| **Atherosclerosis Supplements**** | Peripheral vascular disease | 4.333 |
| **Audiology and Neuro-Otology** | Otorhinolaryngology | 2.318 |

| | | |
|---|---|---|
| **Best Practice & Research In Clinical Rheumatology** | Rheumatology | 3.550 |
| **Biochimica Et Biophysica Acta-Reviews On Cancer** | Oncology | 9.033 |
| **Biological Psychiatry** | Psychiatry | 9.247 |
| **Biology of Blood and Marrow Transplantation** | Transplantation | 3.940 |
| **Biology of Reproduction** | Reproductive biology | 4.027 |
| **Bjog-An International Journal of Obstetrics and Gynaecology** | Obstetrics & Gynecology | 3.760 |
| **Blood** | Hematology | 9.060 |
| **Blood Reviews** | Hematology | 6.000 |
| **Bmc Complementary and Alternative Medicine** | Integrative & complementary medicine | 2.082 |
| **Bmc Family Practice** | Primary health care | 1.609 |
| **Bmc Medicine** | Medicine, general & internal | 6.679 |
| **Bone Marrow Transplantation** | Transplantation | 3.541 |
| **Brain** | Clinical neurology | 9.915 |
| **British Journal of Anaesthesia** | Anesthesiology | 4.237 |
| **British Journal of Dermatology** | Dermatology | 3.759 |
| **British Journal of General Practice** | Primary health care | 2.034 |
| **British Journal of Ophthalmology** | Ophthalmology | 2.725 |
| **British Journal of Oral & Maxillofacial Surgery** | Dentistry, Oral surgery & medicine | 2.717 |
| **British Journal of Psychiatry** | Psychiatry | 6.606 |
| **British Journal of Sports Medicine** | Sport sciences | 3.668 |
| **British Journal of Surgery** | Surgery | 4.839 |
| **British Medical Journal** | Medicine, general & internal | 17.215 |
| **Bulletin of the World Health Organization** | Public, Environmental and Occupational health | 5.250 |
| **Ca-A Cancer Journal For Clinicians** | Oncology | 153.459 |
| **Canadian Family Physician** | Primary health care | 1.808 |
| **Canadian Medical Association Journal** | Medicine, general & internal | 6.465 |
| **Cancer Cell** | Oncology | 24.755 |
| **Cancer Discovery** | Oncology | 10.143 |
| **Cancer Epidemiology Biomarkers & Prevention** | Public, Environmental and Occupational health | 4.559 |
| **Cell Metabolism** | Endocrinology & Metabolism | 14.619 |
| **Cell Transplantation** | Transplantation | 4.422 |
| **Chest** | Critical care medicine; Respiratory system | 5.854 |
| **Circulation** | Cardiac and cardiovasuclar systems; Peripheral vascular disease | 15.202 |

| | | |
|---|---|---|
| **Circulation Research** | Cardiac and cardiovasuclar systems; Hematology; Peripheral vascular disease | 11.861 |
| **Circulation-Cardiovascular Genetics** | Cardiac and cardiovasuclar systems | 6.728 |
| **Circulation-Cardiovascular Imaging** | Radiology, Nuclear medicine and Medical imaging | 5.795 |
| **Circulation-Cardiovascular Interventions** | Cardiac and cardiovasuclar systems | 6.543 |
| **Circulation-Heart Failure** | Cardiac and cardiovasuclar systems | 6.684 |
| **Clinica Chimica Acta** | Medical laboratory technology | 2.850 |
| **Clinical and Experimental Allergy** | Allergy | 4.789 |
| **Clinical Biochemistry** | Medical laboratory technology | 2.450 |
| **Clinical Chemistry** | Medical laboratory technology | 7.149 |
| **Clinical Chemistry and Laboratory Medicine** | Medical laboratory technology | 3.009 |
| **Clinical Gastroenterology and Hepatology** | Gastroenterology & Hepatology | 6.648 |
| **Clinical Implant Dentistry and Related Research** | Dentistry, Oral surgery & medicine | 3.821 |
| **Clinical Infectious Diseases** | Immunology; Infectious diseases | 9.374 |
| **Clinical Journal of the American Society of Nephrology** | Urology & Nephrology | 5.068 |
| **Clinical Oral Implants Research** | Dentistry, Oral surgery & medicine | 3.433 |
| **Clinical Orthopaedics and Related Research** | Orthopedics | 2.787 |
| **Clinical Otolaryngology** | Otorhinolaryngology | 1.869 |
| **Clinical Reviews in Allergy & Immunology** | Allergy | 5.590 |
| **Complementary therapies In Medicine** | Integrative & complementary medicine | 2.093 |
| **Contact Dermatitis** | Allergy; Dermatology | 2.925 |
| **Critical Care** | Critical care medicine | 4.718 |
| **Critical Care Medicine** | Critical care medicine | 6.124 |
| **Critical Reviews In Clinical Laboratory Sciences** | Medical laboratory technology | 3.783 |
| **Current Allergy and Asthma Reports** | Allergy | 2.746 |
| **Current Opinion In Allergy and Clinical Immunology** | Allergy | 3.398 |
| **Current Opinion In Critical Care** | Critical care medicine | 2.967 |
| **Current Opinion In Hiv and Aids** | Infectious diseases | 4.704 |
| **Current Opinion In Immunology** | Immunology | 8.771 |
| **Current Opinion In Infectious Diseases** | Infectious diseases | 4.870 |
| **Current Opinion In Lipidology** | Peripheral vascular disease | 5.839 |
| **Current Opinion In Nephrology and Hypertension** | Urology & Nephrology | 3.964 |
| **Current Opinion In Organ Transplantation** | Transplantation | 3.272 |
| **Current Opinion In Rheumatology** | Rheumatology | 5.191 |
| **Cytometry Part B-Clinical Cytometry** | Medical laboratory technology | 2.231 |
| **Dental Materials** | Dentistry, Oral surgery & medicine | 3.773 |
| **Diabetes** | Endocrinology & Metabolism | 7.895 |
| **Diabetes Care** | Endocrinology & Metabolism | 7.735 |

| | | |
|---|---|---|
| **Dysphagia** | Otorhinolaryngology | 1.938 |
| **Ear and Hearing** | Otorhinolaryngology | 3.262 |
| **Emergencias** | Emergency medicine | 2.578 |
| **Emergency Medicine Journal** | Emergency medicine | 1.645 |
| **Emerging Infectious Diseases** | Infectious diseases | 5.993 |
| **Endocrine Reviews** | Endocrinology & Metabolism | 14.873 |
| **Endoscopy** | Gastroenterology & Hepatology; Surgery | 5.735 |
| **Environmental Health Perspectives** | Public, Environmental and Occupational health | 7.260 |
| **Epidemiologic Reviews** | Public, Environmental and Occupational health | 9.269 |
| **Epidemiology** | Public, Environmental and Occupational health | 5.738 |
| **European Child & Adolescent Psychiatry** | Pediatrics | 3.699 |
| **European Heart Journal** | Cardiac and cardiovasuclar systems | 14.097 |
| **European Journal of Anaesthesiology** | Anesthesiology | 2.792 |
| **European Journal of Epidemiology** | Public, Environmental and Occupational health | 5.118 |
| **European Journal of Nuclear Medicine and Molecular Imaging** | Radiology, Nuclear medicine and Medical imaging | 5.114 |
| **European Journal of Pain** | Anesthesiology | 3.067 |
| **European Respiratory Journal** | Respiratory system | 6.355 |
| **European Urology** | Urology & Nephrology | 10.476 |
| **Eurosurveillance** | Infectious diseases | 5.491 |
| **Evidence-Based Complementary and Alternative Medicine** | Integrative & complementary medicine | 1.722 |
| **Exercise and Sport Sciences Reviews** | Sport sciences | 5.283 |
| **Exercise Immunology Review** | Sport sciences | 7.053 |
| **Experimental Dermatology** | Dermatology | 3.578 |
| **Experimental Eye Research** | Ophthalmology | 3.026 |
| **Experimental Gerontology** | Geriatrics & Gerontology | 3.911 |
| **Family Practice** | Primary health care | 1.828 |
| **Fertility and Sterility** | Obstetrics & Gynecology; Reproductive biology | 4.174 |
| **Frontiers In Aging Neuroscience** | Geriatrics & Gerontology | 5.224 |
| **Frontiers In Neuroendocrinology** | Endocrinology & Metabolism | 7.985 |
| **Gastroenterology** | Gastroenterology & Hepatology | 12.821 |
| **Gastrointestinal Endoscopy** | Gastroenterology & Hepatology | 5.210 |
| **Gut** | Gastroenterology & Hepatology | 10.732 |
| **Gynecologic Oncology** | Obstetrics & Gynecology | 3.929 |
| **Haematologica-the Hematology Journal** | Hematology | 5.935 |
| **Head and Neck-Journal For the Sciences and Specialties of the Head and Neck** | Otorhinolaryngology | 2.833 |
| **Hearing Research** | Otorhinolaryngology | 2.537 |
| **Hepatology** | Gastroenterology & Hepatology | 12.003 |

| | | |
|---|---|---|
| **Human Brain Mapping** | Radiology, Nuclear medicine and Medical imaging | 6.878 |
| **Human Reproduction** | Obstetrics & Gynecology; Reproductive biology | 4.670 |
| **Human Reproduction Update** | Obstetrics & Gynecology; Reproductive biology | 8.847 |
| **Hypertension** | Peripheral vascular disease | 6.873 |
| **Ieee Transactions On Neural Systems and Rehabilitation Engineering** | Rehabilitation | 3.255 |
| **Immunity** | Immunology | 19.795 |
| **Immunological Reviews** | Immunology | 12.155 |
| **Injury-International Journal of the Care of the Injured** | Emergency medicine | 2.174 |
| **Integrative Cancer therapies** | Integrative & complementary medicine | 2.354 |
| **Intensive Care Medicine** | Critical care medicine | 5.258 |
| **International Journal of Epidemiology** | Public, Environmental and Occupational health | 6.982 |
| **International Journal of Oral Science** | Dentistry, Oral surgery & medicine | 2.719 |
| **International Journal of Radiation Oncology Biology Physics** | Radiology, Nuclear medicine and Medical imaging | 4.524 |
| **Investigative Ophthalmology & Visual Science** | Ophthalmology | 3.441 |
| **Investigative Radiology** | Radiology, Nuclear medicine and Medical imaging | 5.460 |
| **Jacc-Cardiovascular Imaging** | Cardiac and cardiovasuclar systems; Radiology, Nuclear medicine and Medical imaging | 6.164 |
| **Jacc-Cardiovascular Interventions** | Cardiac and cardiovasuclar systems | 6.552 |
| **Jaids-Journal of Acquired Immune Deficiency Syndromes** | Infectious diseases | 4.653 |
| **Jama-Journal of the American Medical Association** | Medicine, general & internal | 29.978 |
| **Jaro-Journal of the Association For Research In Otolaryngology** | Otorhinolaryngology | 2.952 |
| **Jnci-Journal of the National Cancer Institute** | Oncology | 14.336 |
| **Journal of Allergy and Clinical Immunology** | Allergy; Immunology | 12.047 |
| **Journal of Alternative and Complementary Medicine** | Integrative & complementary medicine | 1.464 |
| **Journal of Antimicrobial Chemotherapy** | Infectious diseases | 5.338 |
| **Journal of Applied Physiology** | Sport sciences | 3.484 |
| **Journal of Bone and Joint Surgery-American Volume** | Orthopedics | 3.234 |
| **Journal of Cerebral Blood Flow and Metabolism** | Hematology | 5.398 |

| | | |
|---|---|---|
| **Journal of Clinical Epidemiology** | Public, Environmental and Occupational health | 5.332 |
| **Journal of Clinical Oncology** | Oncology | 18.038 |
| **Journal of Clinical Periodontology** | Dentistry, Oral surgery & medicine | 3.688 |
| **Journal of Dental Research** | Dentistry, Oral surgery & medicine | 3.826 |
| **Journal of Dentistry** | Dentistry, Oral surgery & medicine | 3.200 |
| **Journal of Dermatological Science** | Dermatology | 3.520 |
| **Journal of Endodontics** | Dentistry, Oral surgery & medicine | 2.929 |
| **Journal of Ethnopharmacology** | Integrative & complementary medicine | 2.755 |
| **Journal of Experimental Medicine** | Immunology | 13.214 |
| **Journal of Head Trauma Rehabilitation** | Rehabilitation | 4.443 |
| **Journal of Heart and Lung Transplantation** | Respiratory system; Transplantation | 5.112 |
| **Journal of Hepatology** | Gastroenterology & Hepatology | 9.858 |
| **Journal of Infectious Diseases** | Infectious diseases | 5.848 |
| **Journal of Internal Medicine** | Medicine, general & internal | 6.455 |
| **Journal of Investigative Dermatology** | Dermatology | 6.193 |
| **Journal of Mammary Gland Biology and Neoplasia** | Endocrinology & Metabolism | 7.524 |
| **Journal of Manipulative and Physiological therapeutics** | Integrative & complementary medicine | 1.647 |
| **Journal of Neuroengineering and Rehabilitation** | Rehabilitation | 2.567 |
| **Journal of Neurology Neurosurgery and Psychiatry** | Surgery | 4.924 |
| **Journal of Neurotrauma** | Critical care medicine | 4.295 |
| **Journal of Nuclear Medicine** | Radiology, Nuclear medicine and Medical imaging | 5.774 |
| **Journal of Orthopaedic & Sports Physical therapy** | Orthopedics; Rehabilitation; Sport sciences | 2.947 |
| **Journal of Orthopaedic Research** | Orthopedics | 2.875 |
| **Journal of Pediatrics** | Pediatrics | 4.035 |
| **Journal of Physiotherapy** | Rehabilitation | 2.255 |
| **Journal of Pineal Research** | Endocrinology & Metabolism | 7.304 |
| **Journal of Science and Medicine In Sport** | Sport sciences | 2.899 |
| **Journal of the American Academy of Child and Adolescent Psychiatry** | Pediatrics; Psychiatry | 6.970 |
| **Journal of the American Academy of Dermatology** | Dermatology | 4.906 |
| **Journal of the American Board of Family Medicine** | Primary health care | 1.758 |
| **Journal of the American College of Cardiology** | Cardiac and cardiovasuclar systems | 14.086 |
| **Journal of the American College of Surgeons** | Surgery | 4.500 |

| | | |
|---|---|---|
| **Journal of the American Geriatrics Society** | Geriatrics & Gerontology | 3.978 |
| **Journal of the American Medical Directors Association** | Geriatrics & Gerontology | 5.302 |
| **Journal of the American Society of Nephrology** | Urology & Nephrology | 8.987 |
| **Journal of Thoracic and Cardiovascular Surgery** | Respiratory system | 3.526 |
| **Journal of Thoracic Oncology** | Respiratory system | 4.473 |
| **Journal of Thrombosis and Haemostasis** | Hematology; Peripheral vascular disease | 6.081 |
| **Journal of Trauma-Injury Infection and Critical Care** | Emergency medicine | 2.348 |
| **Journal of Tropical Pediatrics** | Tropical medicine | 1.006 |
| **Journal of Urology** | Urology & Nephrology | 3.696 |
| **Journal of Vector Borne Diseases** | Tropical medicine | 1.041 |
| **Journals of Gerontology Series A-Biological Sciences and Medical Sciences** | Geriatrics & Gerontology | 4.314 |
| **Kidney International** | Urology & Nephrology | 7.916 |
| **Lancet** | Medicine, general & internal | 39.060 |
| **Lancet Infectious Diseases** | Infectious diseases | 19.966 |
| **Lancet Neurology** | Clinical neurology | 23.917 |
| **Lancet Oncology** | Oncology | 25.117 |
| **Laryngoscope** | Otorhinolaryngology | 1.979 |
| **Leukemia** | Hematology; Oncology | 10.164 |
| **Liver Transplantation** | Transplantation | 3.944 |
| **Malaria Journal** | Tropical medicine | 3.400 |
| **Medicine and Science In Sports and Exercise** | Sport sciences | 4.475 |
| **Memorias Do Instituto Oswaldo Cruz** | Tropical medicine | 1.363 |
| **Menopause-the Journal of the North American Menopause Society** | Obstetrics & Gynecology | 3.163 |
| **Minerva Anestesiologica** | Anesthesiology; Critical care medicine | 2.818 |
| **Molecular Human Reproduction** | Reproductive biology | 4.542 |
| **Molecular Psychiatry** | Psychiatry | 14.897 |
| **Nature Immunology** | Immunology | 26.199 |
| **Nature Reviews Cancer** | Oncology | 35.000 |
| **Nature Reviews Cardiology** | Cardiac and cardiovasuclar systems | 10.400 |
| **Nature Reviews Clinical Oncology** | Oncology | 15.031 |
| **Nature Reviews Endocrinology** | Endocrinology & Metabolism | 11.025 |
| **Nature Reviews Gastroenterology & Hepatology** | Gastroenterology & Hepatology | 10.426 |
| **Nature Reviews Immunology** | Immunology | 33.129 |
| **Nature Reviews Nephrology** | Urology & Nephrology | 7.943 |
| **Nature Reviews Neurology** | Clinical neurology | 15.518 |
| **Nature Reviews Rheumatology** | Rheumatology | 9.745 |

| | | |
|---|---|---|
| **Nature Reviews Urology** | Urology & Nephrology | 4.793 |
| **Nephrology Dialysis Transplantation** | Transplantation | 3.371 |
| **Neurobiology of Aging** | Geriatrics & Gerontology | 6.166 |
| **Neurocritical Care** | Critical care medicine | 3.038 |
| **Neuroimage** | Radiology, Nuclear medicine and Medical imaging | 6.252 |
| **Neurology** | Clinical neurology | 8.249 |
| **Neuro-Oncology** | Clinical neurology | 6.180 |
| **Neuropsychopharmacology** | Psychiatry | 8.678 |
| **Neurorehabilitation and Neural Repair** | Rehabilitation | 4.278 |
| **New England Journal of Medicine** | Medicine, general & internal | 51.658 |
| **Obstetrics and Gynecology** | Obstetrics & Gynecology | 4.798 |
| **Ocular Surface** | Ophthalmology | 2.643 |
| **Ophthalmology** | Ophthalmology | 5.563 |
| **Osteoarthritis and Cartilage** | Orthopedics; Rheumatology | 4.262 |
| **Otology & Neurotology** | Otorhinolaryngology | 2.014 |
| **Pain** | Anesthesiology | 5.644 |
| **Pain Practice** | Anesthesiology | 2.605 |
| **Pediatric Allergy and Immunology** | Allergy; Pediatrics | 3.376 |
| **Pediatric Infectious Disease Journal** | Pediatrics | 3.569 |
| **Pediatrics** | Pediatrics | 5.119 |
| **Periodontology 2000** | Dentistry, Oral surgery & medicine | 4.012 |
| **Physical therapy** | Orthopedics; Rehabilitation | 2.778 |
| **Phytomedicine** | Integrative & complementary medicine | 2.972 |
| **Pigment Cell & Melanoma Research** | Dermatology | 5.839 |
| **Placenta** | Reproductive biology | 3.117 |
| **Plos Medicine** | Medicine, general & internal | 15.253 |
| **Plos Neglected Tropical Diseases** | Tropical medicine | 4.569 |
| **Prehospital Emergency Care** | Emergency medicine | 1.859 |
| **Primary Care Diabetes** | Primary health care | 1.609 |
| **Primary Care Respiratory Journal** | Primary health care | 2.191 |
| **Progress In Retinal and Eye Research** | Ophthalmology | 9.439 |
| **Prostate** | Urology & Nephrology | 3.843 |
| **Psychotherapy and Psychosomatics** | Psychiatry | 7.230 |
| **Radiology** | Radiology, Nuclear medicine and Medical imaging | 6.339 |
| **Radiotherapy and Oncology** | Radiology, Nuclear medicine and Medical imaging | 4.520 |
| **Regional Anesthesia and Pain Medicine** | Anesthesiology | 3.464 |
| **Reproduction** | Reproductive biology | 3.555 |
| **Reproductive Toxicology** | Reproductive biology | 3.141 |
| **Respiratory Research** | Respiratory system | 3.642 |
| **Resuscitation** | Critical care medicine; Emergency medicine | 4.104 |

| | | |
|---|---|---|
| **Retina-the Journal of Retinal and Vitreous Diseases** | Ophthalmology | 2.825 |
| **Rheumatology** | Rheumatology | 4.212 |
| **Scandinavian Journal of Medicine & Science In Sports** | Sport sciences | 3.214 |
| **Scandinavian Journal of Primary Health Care** | Primary health care | 1.905 |
| **Scandinavian Journal of Trauma Resuscitation & Emergency Medicine** | Emergency medicine | 1.680 |
| **Schizophrenia Bulletin** | Psychiatry | 8.486 |
| **Seminars In Arthritis and Rheumatism** | Rheumatology | 3.806 |
| **Seminars In Fetal & Neonatal Medicine** | Pediatrics | 3.505 |
| **Seminars In Liver Disease** | Gastroenterology & Hepatology | 8.274 |
| **Seminars In Reproductive Medicine** | Obstetrics & Gynecology; Reproductive biology | 3.211 |
| **Seminars In Thrombosis and Hemostasis** | Peripheral vascular disease | 4.216 |
| **Skin Pharmacology and Physiology** | Dermatology | 2.885 |
| **Sleep Medicine Reviews** | Clinical neurology | 8.681 |
| **Spine Journal** | Orthopedics | 3.220 |
| **Sports Medicine** | Sport sciences | 5.237 |
| **Stem Cells** | Hematology | 7.701 |
| **Stem Cells and Development** | Transplantation | 4.670 |
| **Stroke** | Peripheral vascular disease | 6.158 |
| **Supportive Care In Cancer** | Rehabilitation | 2.649 |
| **Surgery For Obesity and Related Diseases** | Surgery | 4.121 |
| **Survey of Ophthalmology** | Ophthalmology | 2.859 |
| **Therapeutic Drug Monitoring** | Medical laboratory technology | 2.234 |
| **Thorax** | Respiratory system | 8.376 |
| **Thrombosis and Haemostasis** | Hematology; Peripheral vascular disease | 6.094 |
| **Transactions of the Royal Society of Tropical Medicine and Hygiene** | Tropical medicine | 1.823 |
| **Translational Research** | Medical laboratory technology | 3.490 |
| **Transplantation** | Transplantation | 3.781 |
| **Trends In Endocrinology and Metabolism** | Endocrinology & Metabolism | 8.901 |
| **Trends In Immunology** | Immunology | 9.486 |
| **Tropical Medicine & International Health** | Tropical medicine | 2.938 |
| **Ultrasound In Obstetrics & Gynecology** | Obstetrics & Gynecology | 3.557 |
| **Who Technical Report Series**** | Public, Environmental and Occupational health | 6.100 |
| **World Psychiatry** | Psychiatry | 8.974 |

*Subject category 2012 Journal Citation Reports ®)[1], some journals belong to more than one category

** Excluded for analyses, no journal website with up-to-date information identified

**Supplemental Table 2. Reporting guidelines mentioned in the Instructions to Authors on journals' websites (n=337), ranked based on the 2018 results**

| Reporting guideline | 2017 | 2018 |
|---|---|---|
| CONSORT (COnsolidated Standards Of Reporting Trials; http://www.consort-statement.org/)[2,3] | 170 (50%) | 178 (53%) |
| PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; http://www.prisma-statement.org/)[4,5] | 115 (34%) | 141 (42%) |
| STROBE (Strengthening the Reporting of Observational Studies in Epidemiology; https://www.equator-network.org/reporting-guidelines/strobe/)[6,7] | 88 (26%) | 107 (32%) |
| ARRIVE (Animal Research: Reporting of In Vivo Experiments; https://www.nc3rs.org.uk/arrive-guidelines)[8] | 80 (24%) | 95 (28%) |
| STARD (Standards for Reporting Diagnostic accuracy studies; https://www.equator-network.org/reporting-guidelines/stard/)[9,10] | 82 (24%) | 92 (27%) |
| MOOSE (Meta-analysis Of Observational Studies in Epidemiology; https://www.equator-network.org/reporting-guidelines/meta-analysis-of-observational-studies-in-epidemiology-a-proposal-for-reporting-meta-analysis-of-observational-studies-in-epidemiology-moose-group/)[11] | 52 (15%) | 53 (16%) |
| CONSORT-Extensions* (COnsolidated Standards Of Reporting Trials; http://www.consort-statement.org/) | 36 (11%) | 38 (11%) |
| - *STRICTA (Standards for Reporting Interventions in Clinical Trials of Acupuncture; https://www.stricta.info/)[12]* | *2* | *2* |
| - *RedHot (Reporting data on homeopathic treatments)[13]* | *1* | *1* |
| - *Not specified* | *33* | *35* |
| CHEERS (Consolidated Health Economic Evaluation Reporting Standards; https://www.equator-network.org/reporting-guidelines/cheers/)[14] | 25 (7%) | 36 11%) |
| TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis; http://www.tripod-statement.org/)[15,16] | 27 (8%) | 29 (9%) |
| STROBE-Extensions* (Strengthening the Reporting of Observational Studies in Epidemiology; https://www.equator-network.org/reporting-guidelines/strobe/)[6,7] | 2 (1%) | 3 (1%) |
| - *STREGA (STrengthening the REporting of Genetic Association Studies)[17]* | *18* | *17* |
| - *RECORD (REporting of studies Conducted using Observational Routinely-collected Data; https://www.record-statement.org/)[18]* | *6* | *5* |
| - *STROME-ID (Strengthening the reporting of molecular epidemiology for infectious diseases)[19]* | *1* | *1* |
| - *Not specified* | *2* | *3* |
| CARE (CAse REport guidelines; https://www.care-statement.org/)[20] | 16 (5%) | 26 (8%) |

| | | |
|---|---|---|
| COREQ (Consolidated criteria for reporting qualitative research; https://www.equator-network.org/reporting-guidelines/coreq/)[21] | 18 (5%) | 24 (7%) |
| SQUIRE (Standards for QUality Improvement Reporting Excellence; http://www.squire-statement.org/)[22] | 18 (5%) | 22 (7%) |
| SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials; http://www.spirit-statement.org/)[23] | 14 (4%) | 22 (7%) |
| PRISMA-Extensions* (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; http://www.prisma-statement.org/) | 12 (4%) | 15 (4%) |
| REMARK (REporting recommendations for tumour MARKer prognostic studies; https://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tumour-marker-prognostic-studies-remark/)[24,25] | 12 (4%) | 14 (4%) |
| MIAME (Minimum Information About a Microarray Experiment)[26] | 0 (0%) | 13 (4%) |
| SRQR (Standards for reporting qualitative research; https://www.equator-network.org/reporting-guidelines/srqr/)[27] | 4 (1%) | 13 (4%) |
| SAMPL (Statistical Analyses and Methods in the Published Literature; https://www.equator-network.org/reporting-guidelines/sampl/)[28] | 13 (4%) | 13 (4%) |
| TREND (Transparent Reporting of Evaluations with Nonrandomized Designs; https://www.cdc.gov/trendstatement/)[29] | 10 (3%) | 11 (3%) |
| AGREE (Appraisal of Guidelines, REsearch and Evaluation; https://www.agreetrust.org/resource-centre/agree-reporting-checklist/)[30] | 0 (0%) | 7 (2%) |
| ENTREQ (Enhancing transparency in reporting the synthesis of qualitative research; https://www.equator-network.org/reporting-guidelines/entreq/)[31] | 4 (1%) | 6 (2%) |
| TIDieR (Template for Intervention Description and Replication; http://www.tidierguide.org/)[32] | 3 (1%) | 6 (2%) |
| NIH (Principles and Guidelines for Reporting Preclinical Research - National Insitute of Health; https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research) | 0 (0%) | 6 (2%) |
| GATHER (Guidelines for Accurate and Transparent Health Estimates Reporting; http://gather-statement.org/)[33] | 4 (1%) | 4 (1%) |
| QUORUM (Quality of Reporting of Meta-analyses standards); replaced by PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; http://www.prisma-statement.org/)[4,5] | 4 (1%) | 3 (1%) |
| BRISQ (Biospecimen Reporting for Improved Study Quality; https://www.equator-network.org/reporting-guidelines/brisq/)[34] | 3 (1%) | 3 (1%) |
| GRIPS (Strengthening the reporting of Genetic RIsk Prediction Studies; https://www.equator-network.org/reporting-guidelines/strengthening-the-reporting-of-genetic-risk-prediction-studies-the-grips-statement/)[35] | 3 (1%) | 3 (1%) |

| | | |
|---|---|---|
| GRRAS (Guidelines for Reporting Reliability and Agreement Studies; https://www.equator-network.org/reporting-guidelines/guidelines-for-reporting-reliability-and-agreement-studies-grras-were-proposed/)[36] | 3 (1%) | 3 (1%) |
| ORION (Guidelines for transparent reporting of Outbreak Reports and Intervention studies Of Nosocomial infection; https://www.ucl.ac.uk/amr/Reporting_Guidelines/ORION)[37] | 2 (1%) | 2 (1%) |
| CHERRIES (Checklist for Reporting Results of Internet E-Surveys; https://www.equator-network.org/reporting-guidelines/improving-the-quality-of-web-surveys-the-checklist-for-reporting-results-of-internet-e-surveys-cherries/)[38] | 1 (0.3%) | 2 (1%) |
| MIQE (minimum information for publication of quantitative real-time PCR experiments)[39] | 0 (0%) | 2 (1%) |
| STROND (Standards of Reporting of Neurological Disorders; https://www.equator-network.org/reporting-guidelines/development-of-the-standards-of-reporting-of-neurological-disorders-strond-checklist-a-guideline-for-the-reporting-of-incidence-and-prevalence-studies-in-neuroepidemiology/)[40] | 1 (0.3%) | 1 (0.3%) |
| GNOSIS (guidelines for neuro-oncology: standards for investigational studies reporting of phase 1 and phase 2 clinical trials; https://www.equator-network.org/reporting-guidelines/gnosis-guidelines-for-neuro-oncology-standards-for-investigational-studies-reporting-of-phase-1-and-phase-2-clinical-trials/)[41] | 1 (0.3%) | 1 (0.3%) |
| HuGENet (https://www.equator-network.org/reporting-guidelines/the-hugenet-huge-review-handbook-version-1-0-guidelines-for-systematic-review-and-meta-analysis-of-gene-disease-association-studies/)[42] | 1 (0.3%) | 1 (0.3%) |
| STRENDA (Standards for Reporting Enzymology Data; https://www.beilstein-strenda-db.org/strenda/)[43] | 1 (0.3%) | 1 (0.3%) |
| SCRIBE (Single-Case Reporting Guideline In BEhavioural Interventions; http://sydney.edu.au/medicine/research/scribe/)[44] | 1 (0.3%) | 1 (0.3%) |
| RAMESES (http://www.ramesesproject.org/)[45] | 1 (0.3%) | 1 (0.3%) |
| COS-STAR (Core Outcome Set-STAndards for Reporting; https://www.equator-network.org/reporting-guidelines/cos-star-statement/)[46] | 0 (0%) | 1 (0.3%) |
| STARi (Standards for Reporting Implementation Studies; https://www.equator-network.org/reporting-guidelines/stari-statement/)[47,48] | 0 (0%) | 1 (0.3%) |

Number of journals (%)
*Numbers present any extension of the reporting guideline mentioned.

**Supplemental Table 3. Journal characteristics of journals mentioning TRIPOD in their instructions to authors**

| | 2017 (n=27 journals) | 2018 (n=29 journals) |
|---|---|---|
| **Journal Impact factor** | 3.6 ($P_{25}$–$P_{75}$: 2.6-6.1) | 3.7 ($P_{25}$–$P_{75}$: 2.6-6.1) |
| **Clinical domains*** | | |
| Anesthesiology | 1 | 1 |
| Critical care medicine | 1 | 2 |
| Dermatology | 1 | 1 |
| Emergency medicine | 2 | 1 |
| Gastroenterology | 1 | 1 |
| Hematology | 1 | 2 |
| Integrative & complementary Medicine | 1 | 1 |
| Medical, general and Internal Medicine | 2 | 3 |
| Oncology | 1 | 1 |
| Ophthalmology | 1 | 0 |
| Orthopedics | 1 | 1 |
| Pediatrics | 2 | 2 |
| Peripheral vascular disease | 1 | 2 |
| Primary health care | 2 | 2 |
| Rehabilitation | 3 | 4 |
| Respiratory system | 2 | 1 |
| Rheumatology | 2 | 1 |
| Sport sciences | 1 | 2 |
| Surgery | 1 | 0 |
| Transplantation | 1 | 1 |
| Tropical medicine | 1 | 1 |
| Urology & Nephrology | 1 | 3 |

*Numbers add up to over 27 and 29, respectively, as a journal could belong to more than one category.

**Supplemental Table 4. Summary of survey responses on journal and respondent characteristics**

| | Number of respondents | |
|---|---|---|
| **Type of journal** | **52** | |
| General | | 10 (19%) |
| Specialized | | 39 (75%) |
| Unknown | | 3 (6%) |
| **Journal Impact factor** | **49** | 4.3 (2.8-6.9) |
| **Clinical domains*** | **52** | |
| Allergy | | 2 (4%) |
| Anesthesiology | | 3 (6%) |
| Clinical Neurology | | 4 (8%) |
| Critical care medicine | | 1 (2%) |
| Dermatology | | 1 (2%) |
| Emergency medidine | | 2 (4%) |
| Gastroenterology | | 1 (2%) |
| Geriatrics & gerontology | | 1 (2%) |
| Hematology | | 1 (2%) |
| Infectious disease | | 2 (4%) |
| Medical General and Internal Medicine | | 2 (4%) |
| Medical Laboratory technology | | 2 (4%) |
| Oncology | | 2 (4%) |
| Ophthalmology | | 1 (2%) |
| Orthopedics | | 3 (6%) |
| Otorhinolaringology | | 2 (4%) |
| Pediatrics | | 3 (6%) |
| Peripheral vascular disease | | 2 (4%) |
| Psychiatry | | 1 (2%) |
| Public, Environmental and Occupational health | | 4 (8%) |
| Radiology, Nuclear medicine and Medical imaging | | 1 (2%) |
| Rehabilitation | | 2 (4%) |
| Rheumatology | | 1 (2%) |
| Transplantation | | 1 (2%) |
| Tropical medicine | | 3 (6%) |
| Urology & Nephrology | | 3 (6%) |
| Unknown | | 4 (8%) |
| **Study designs published*** | **49** | |
| Systematic reviews | | 45 (92%) |
| Observational studies | | 40 (82%) |
| Randomised trials | | 37 (76%) |
| Diagnostic or prognostic studies | | 36 (73%) |
| Qualitative research | | 29 (59%) |

| | | |
|---|---|---|
| Quality improvement studies | | 26 (53%) |
| Economic evaluations | | 24 (49%) |
| Animal pre-clinical studies | | 24 (49%) |
| Case reports | | 18 (37%) |
| Study protocols | | 4 (8%) |
| (Narrative) reviews and opinion pieces | | 2 (4%) |
| Clinical practice guidelines | | 1 (2%) |
| **Respondents** | **52** | |
| Editor-in-chief | | 37 (71%) |
| Managing editor | | 7 (13%) |
| Editor | | 5 (10%) |
| Unknown | | 3 (6%) |
| **Familiar with reporting guidelines** | **50** | 44 (88%) |
| **Familiar with the EQUATOR Network** | **42** | 34 (81%) |
| **Familiarity with specific reporting guidelines** | **41** | |
| CONSORT | | 37 (90%) |
| PRISMA | | 36 (88%) |
| STROBE | | 36 (88%) |
| STARD | | 21 (51%) |
| TRIPOD | | 18 (44%) |
| ARRIVE | | 13 (32%) |
| SQUIRE | | 11 (27%) |
| CARE | | 9 (22%) |
| CHEERS | | 9 (22%) |
| SPIRIT | | 8 (20%) |
| COREQ | | 5 (12%) |
| SRQR | | 2 (5%) |
| Other | | ** |
| **Learn about new reporting guidelines or extensions to existing reporting guidelines through*** | **45** | |
| The EQUATOR network | | 26 (58%) |
| Colleagues | | 25 (56%) |
| Publications about (development of) reporting guidelines | | 20 (44%) |
| Conference presentations | | 14 (31%) |
| Authors | | 14 (31%) |
| Other | | 4 (9%)*** |
| **Endorsing reporting guidelines** | **41** | **35 (85%)** |
| **Specific reporting guidelines endorsed** | **34** | |
| CONSORT | | 30 (88%) |
| STROBE | | 26 (76%) |
| PRISMA | | 25 (74%) |
| STARD | | 12 (35%) |
| ARRIVE | | 11 (32%) |
| TRIPOD | | 8 (24%) |

| | |
|---|---|
| SQUIRE | 6 (18%) |
| CARE | 4 (12%) |
| COREQ | 4 (12%) |
| CHEERS | 3 (9%) |
| SPIRIT | 1 (3%) |
| SRQR | 0 (0%) |
| Other | **** |

n(%); median ($P_{25}$–$P_{75}$)
*Adds up to over 100%, as more than one category could be applicable
**Other reporting guidelines mentioned by respondents: ISPOR (1), MIAME (1), MIQE (1), MOOSE (1), REMARK (1), ORION (1), QHES (1), RECORD (1), SAMPL (1), STREGA(2), TREND (1)
*** Answers provided: Council of Science Editors (1), journal work / editorial meetings (2), as manuscript submitted for publication to own journal (1).
****Other reporting guidelines mentioned by respondents: STREGA (3), TREND (2), MOOSE (2), MIAME (2), AGREE (1), ISPOR (1), MIQE (1), QHES (1), ORION (1), RECORD (1), REMARK (1), SAGER (1), SAMPL (1); in addition EQUATOR (4), ICMJE (2), Resource identification initiative (1) were mentioned and one journal stated to endorse all reporting guidelines that exist.

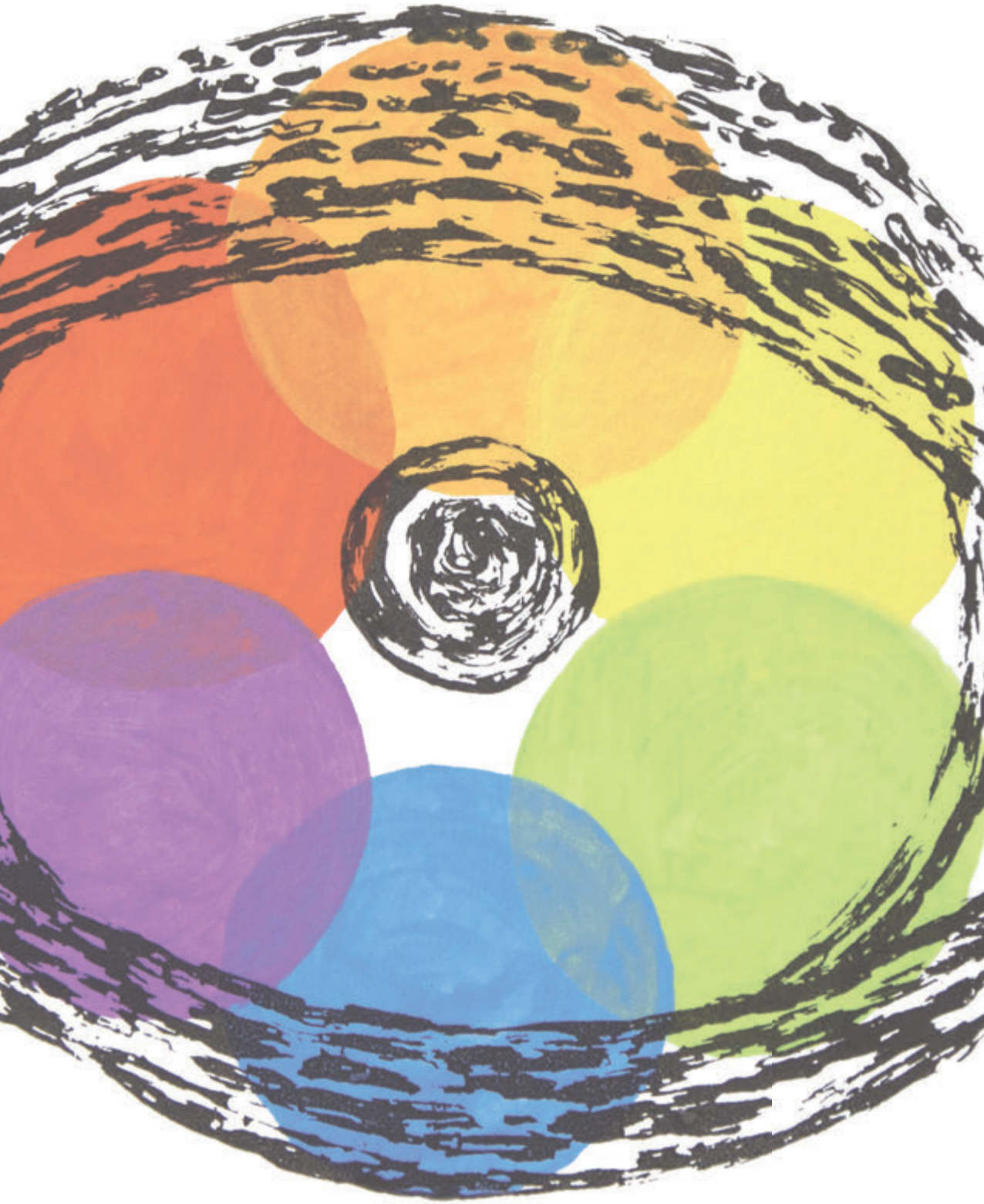
**References**


1. 2012 Journal Citation Reports® In. Science Edition ed: Clarivate Analytics 2017.
2. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.
3. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152(11):726-732.
4. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med*. 2009;6(7):e1000100.
5. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
6. Vandenbroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med*. 2007;4(10):e297.
7. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573-577.
8. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol*. 2010;8(6):e1000412.
9. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
10. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.
11. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283(15):2008-2012.
12. MacPherson H, Altman DG, Hammerschlag R, Youping L, Taixiang W, et al. Revised STandards for Reporting Interventions in Clinical Trials of Acupuncture (STRICTA): extending the CONSORT statement. *PLoS Med*. 2010;7(6):e1000261.
13. Dean ME, Coulter MK, Fisher P, Jobst K, Walach H. Reporting data on homeopathic treatments (RedHot): a supplement to CONSORT*. *Forsch Komplementmed*. 2006;13(6):368-371.

14. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Value Health*. 2013;16(2):e1-5.
15. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63.
16. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-73.
17. Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, et al. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. *PLoS Med*. 2009;6(2):e22.
18. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12(10):e1001885.
19. Field N, Cohen T, Struelens MJ, Palm D, Cookson B, et al. Strengthening the Reporting of Molecular Epidemiology for Infectious Diseases (STROME-ID): an extension of the STROBE statement. *Lancet Infect Dis*. 2014;14(4):341-352.
20. Gagnier JJ, Kienle G, Altman DG, Moher D, Sox H, Riley D. The CARE Guidelines: Consensus-based Clinical Case Reporting Guideline Development. *Glob Adv Health Med*. 2013;2(5):38-43.
21. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007;19(6):349-357.
22. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf*. 2016;25(12):986-992.
23. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gotzsche PC, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med*. 2013;158(3):200-207.
24. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer*. 2005;93(4):387-391.
25. Sauerbrei W, Taube SE, McShane LM, Cavenagh MM, Altman DG. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): An Abridged Explanation and Elaboration. *J Natl Cancer Inst*. 2018;110(8):803-811.
26. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*. 2001;29(4):365-371.

27. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med*. 2014;89(9):1245-1251.
28. Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines. *Int J Nurs Stud*. 2015;52(1):5-9.
29. Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*. 2004;94(3):361-366.
30. Brouwers MC, Kerkvliet K, Spithoff K. The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. *BMJ*. 2016;352:i1152.
31. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med Res Methodol*. 2012;12:181.
32. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014;348:g1687.
33. Stevens GA, Alkema L, Black RE, Boerma JT, Collins GS, et al. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *Lancet*. 2016;388(10062):e19-e23.
34. Moore HM, Kelly AB, Jewell SD, McShane LM, Clark DP, et al. Biospecimen reporting for improved study quality (BRISQ). *Cancer Cytopathol*. 2011;119(2):92-101.
35. Janssens AC, Ioannidis JP, van Duijn CM, Little J, Khoury MJ. Strengthening the reporting of Genetic RIsk Prediction Studies: the GRIPS Statement. *PLoS Med*. 2011;8(3):e1000420.
36. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64(1):96-106.
37. Stone SP, Cooper BS, Kibbler CC, Cookson BD, Roberts JA, et al. The ORION statement: guidelines for transparent reporting of outbreak reports and intervention studies of nosocomial infection. *J Antimicrob Chemother*. 2007;59(5):833-840.
38. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res*. 2004;6(3):e34.
39. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem*. 2009;55(4):611-622.

40. Bennett DA, Brayne C, Feigin VL, Barker-Collo S, Brainin M, et al. Development of the Standards of Reporting of Neurological Disorders (STROND) checklist: A guideline for the reporting of incidence and prevalence studies in neuroepidemiology. *Neurology*. 2015;85(9):821-828.
41. Chang SM, Reynolds SL, Butowski N, Lamborn KR, Buckner JC, et al. GNOSIS: guidelines for neuro-oncology: standards for investigational studies-reporting of phase 1 and phase 2 clinical trials. *Neuro Oncol*. 2005;7(4):425-434.
42. *The HuGENet™ HuGE Review Handbook, version 1.0*. Centers for Disease Control and Prevention; 2006.
43. Swainston N, Baici A, Bakker BM, Cornish-Bowden A, Fitzpatrick PF, et al. STRENDA DB: enabling the validation and sharing of enzyme kinetics data. *Febs j*. 2018;285(12):2193-2204.
44. Tate RL, Perdices M, Rosenkoetter U, Shadish W, Vohra S, et al. The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 Statement. *J Clin Epidemiol*. 2016;73:142-152.
45. Wong G, Westhorp G, Manzano A, Greenhalgh J, Jagosh J, Greenhalgh T. RAMESES II reporting standards for realist evaluations. *BMC Med*. 2016;14(1):96.
46. Kirkham JJ, Gorst S, Altman DG, Blazeby JM, Clarke M, et al. Core Outcome Set-STAndards for Reporting: The COS-STAR Statement. *PLoS Med*. 2016;13(10):e1002148.
47. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, et al. Standards for Reporting Implementation Studies (StaRI) Statement. *BMJ*. 2017;356:i6795.
48. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, et al. Standards for Reporting Implementation Studies (StaRI): explanation and elaboration document. *BMJ Open*. 2017;7(4):e013318.



# Chapter 6

## Barriers and facilitators to reduce low-value care: a qualitative evidence synthesis

Simone A van Dulmen*
Christiana A Naaktgeboren*
Pauline Heus
Eva W Verkerk
Jan-Willem Weenink
Rudolf B Kool
Lotty Hooft

*Authors contributed equally

*Submitted*

## Abstract

### Background

The need to reduce costs and harms associated with low-value care practices are increasingly receiving attention. Insight into factors that hamper or facilitate the reduction of low-value care facilitates the design of effective de-implementation strategies. This review aims to assess barriers and facilitators to de-implementation.

### Methods

A qualitative evidence synthesis was performed with a framework analysis. Medline, Embase, Cochrane Library, and Rx for Change databases from 1990 until September 2018 were searched.

### Results

We identified 404 factors in 111 articles. 55% were classified as barriers, 18% as facilitators, 9% as both barrier and facilitator. 18% were factors identified based on articles that measured the effect of a de-implementation strategy; these could not be classified as a barrier or facilitator. Factors related to the individual provider (n=131) were associated with their attitude (n=72; 55%), knowledge/skills (n=43; 33%), behavior (n=11; 8%), and provider characteristics (n=5; 4%). Individual patient factors (n=58) were mainly related to knowledge (n=33; 56%) and attitude (n=13; 22%). Factors related to the social context (n=46) included mainly professional teams (n=23; 50%) and professional development (n=12; 26%). Frequent factors in the organizational context (n=67) were available resources (n=28; 41%) and organizational structures and work routines (n=24; 36%). Under the category of economic and political context (n=31), financial incentives were most common (n=27; 87%).

### Conclusions

Insights into barriers and facilitators to de-implementation provided by this evidence synthesis can improve the design and execution of de-implementation strategies. As most studies found factors on multiple levels, we conclude that multifaceted de-implementation strategies are often necessary for effective reduction of low-value care. Situation-specific knowledge of impeding or facilitating factors across all levels is important for designing tailored de-implementation strategies.

## Introduction

Healthcare with no or little benefit for the patient given the available alternatives, costs and preferences, is an increasingly recognized problem that affects costs, patient safety and satisfaction.[1,2] Several recent initiatives identified such low-value care practices, including the NICE do-not-do list and Choosing Wisely.[3-6] However, simply identifying low-value care is not sufficient for its abandonment.[7,8]

The active process of reducing low-value care has various names such as de-adoption, disinvestment, or de-implementation.[9,10] While de-implementation has several parallels to implementation, many have argued that stopping or changing an existing practice is likely to be more difficult than starting a new one.[11-16] Interventions to reduce low-value care should be targeted at the factors influencing de-implementation or the continuation of low-value care.

Increasing our understanding of the active process of de-implementing low-value care will help such interventions to become more efficient and sustainable manner. Recent reviews have described the effectiveness of interventions to reduce low-value care and the current approaches and challenges to such processes.[9,17] For example, a review by Colla et al. found that effectiveness of strategies varied widely and concluded that it is important to consider the context of the system in which the intervention is implemented.[17] A scoping review by Niven et al. identified knowledge gaps in the field and pointed to the need for a systematic exploration of the barriers and facilitators to de-implementation of low-value care.[9] In their framework, they classified facilitators and barriers to de-adoption of low-value care, as many experts consider this as a key step prior to designing and tailoring an effective de-implementation strategy. Niven et al. concluded that a systematic exploration of the barriers and facilitators to de-implementation of low-value care is an important knowledge gap.[9]

The aim of our qualitative evidence synthesis is to identify and categorize the existing evidence on barriers and facilitators for de-implementation of low-value care. The results of this overview contribute to the knowledge base on de-implementation and might create awareness on the identification of barriers and facilitators for de-implementation. This can be used by healthcare professionals and researchers in developing tailor-made de-implementation strategies aimed at reducing low-value care.

## Methods

### Study design and search strategy

A qualitative evidence synthesis was performed with a framework analysis,[18,19] based on a predefined framework developed by Grol and Wensing for grouping barriers and facilitators for change.[20] The synthesis included articles that identified barriers and facilitators for de-implementation of low-value care. We performed a systematic search to identify relevant studies in using synonyms for de-implementation and low-value care. The search was run in Embase, Medline, and Rx for Change databases on 12th September, 2018. Websites of healthcare quality improvement organizations were also searched and reference checking was performed. Details of the search strategy can be found in the Supplemental Appendix.

### Study selection

We included articles published in English, German, French, or Dutch published after 1990 that identified barriers or facilitators for de-implementation or the presence of low-value care in an original study. Studies that primarily focused on identifying factors influencing de-implementation or the continuation of low-value care were included. We also included studies evaluating the effect of a de-implementation strategy, in which determinants related to the effect of the intervention were measured (evidence-based factors), or in which the authors reflected on potential barriers and facilitators related to the effect of the intervention, e.g. in the discussion section (expert-based factors). For protocols and conference abstracts, we checked whether the study had been published as a full text. Articles on guideline adherence were only included when the aim of the study was explicitly stated as reducing low-value healthcare practices. Articles on disinvestment, in which the motivation for reduction or removal is primarily financial, were excluded. Review articles were also excluded because they have often a broader scope than factors related to de-implementation of low-value care.

Any type of care practice was eligible, including diagnostic and therapeutic practices. No judgment was made whether the particular test or treatment was indeed of low-value; we relied on authors' statements.

Titles and abstracts were screened by two authors and for selected articles, eligibility was based on full text and judged by two authors (C.A.N., J.W., P.H., E.V., L.H., and S.D.). A third author was consulted to resolve discrepancies.

### Data extraction

Data extraction was performed by one author and a second author was consulted when there were doubts (C.A.N., J.W., P.H., E.V., S.D. and L.H.). We used a pre-designed electronic form that was pilot tested using a random sample of 15 articles by all reviewers. Uncertainties or difficulties in data extraction were discussed during face-to-face sessions to ensure consistent extraction of the data.

### Categorization of factors

The factors were classified based on a framework developed by Grol and Wensing,[20,21] which contains five levels: individual provider, individual patient, social context, organizational context, economic and political context. The levels of individual provider and patient are divided in three subcategories: knowledge and skills, attitudes, and behavioral factors and routines.[21] The category social context is divided in professional development, professional teams, and professional networks. The level of organizational context consist of three subcategories; structures and work routines, organizational processes, and available resources. The economic and political context is divided in financial incentives, legal regulatory measures, and segment of target groups.



If possible, we distinguished barriers from facilitators. Many factors were explicitly described as a barrier or a facilitator. An example of a barrier is when providers indicate 'that their time with the patient is too limited to talk to them about the merits of the treatment plan or what options they have'.[22] In some cases, however, it was not clear from wording whether a factor was perceived as a barrier or a facilitator. For example, one article reported that 'multidisciplinary structure of teams and quality of interaction among group members are factors related to de-implementation'.[23] These were categorized as 'both a barrier and a facilitator'. Factors that ware identified based on the articles that measured the effect of a de-implementation strategy (e.g. in subgroup or multivariable regression analyses) were classified in an additional category as 'determinants'. Determinants may often be a proxy for factors related to, for example, knowledge or behaviors of patients and providers, as they may not be directly barriers and facilitators in and of themselves. Therefore, we analyzed them separately. The results are reported, in so far as relevant, according to the guidelines for reporting a synthesis of qualitative research, the ENTREQ guidelines.[24]

## Results

### Search results

The search resulted in 4111 titles and abstracts to screen. After exclusion of 3451 articles based on title and abstract screening, 660 articles were full text screened, of

which 111 were included. Details of the search and selection process are presented in Figure 1 and a list of the included articles can be found in the Supplemental Appendix.

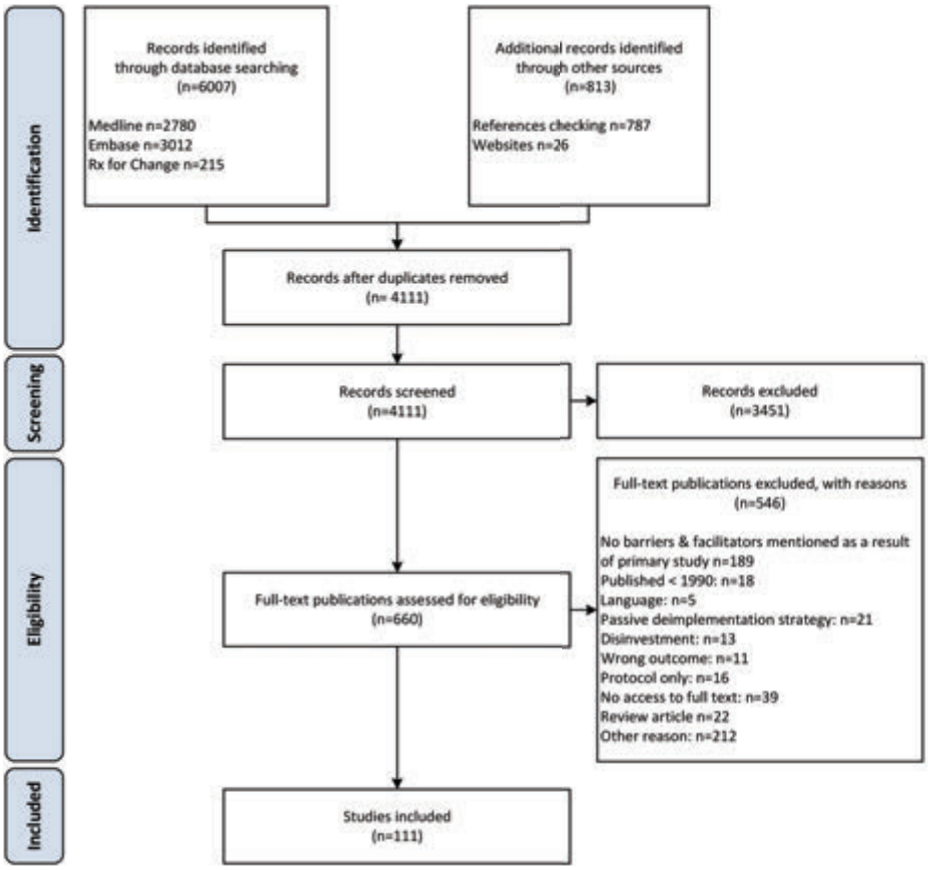



**Figure 1. Study flow**

### Characteristics of included articles

In terms of study design, the majority of the articles were quantitative studies (n=60; 54%). Seven randomized controlled trials (RCTs) were found; the others were observational studies, most without a parallel control arm. 34% (n=38) of the articles had a qualitative component: only interviews or focus groups (n=23), or a combination of quantitative and qualitative methods (n=13).

The study characteristics of the included articles are described in Table 1. The primary aim of 62 (56%) articles was to identify factors influencing de-implementation or the continuation of low-value care, and 49 (44%) aimed to evaluate the effectiveness

of a de-implementation strategy. The majority of the articles (n=88; 80%) focused on therapeutic low-value care practices. Antibiotics (n=39) were by far the most commonly studied therapeutic practice, followed by gastric acid suppressants (n=10). Of the articles that focused on diagnostic tests (n=39), imaging and laboratory tests were the most studied (n=14 and n=12 respectively).

**Table 1. Characteristics of included articles (n=111)**

| | **Studies n (%)** |
|---|---|
| **Study design** | |
| Randomized controlled trial | 7 (6) |
| Non randomized controlled trial | 10 (9) |
| Before after design / interrupted time series | 26 (23) |
| Cohort study | 12 (11) |
| Chart review | 5 (5) |
| Qualitative research design | 23 (21) |
| Survey | 15 (14) |
| Mixed methods | 13 (12) |
| **Aim of the article** | |
| Identify factors influencing de-implementation or the continuation of low-value care | 62 (56) |
| Measure the effectiveness of de-implementation | 49 (44) |
| **Low-value care practice under study*** | |
| **Therapeutic** | **83 (75)** |
| Drug | 59 (53) |
| Antibiotic | 39 (35) |
| Gastric acid suppressants | 10 (9) |
| Polypharmacy | 5 (5) |
| Benzodiazepine, opioids, analgesic, psychotropic | 5 (5) |
| Blood or albumin transfusion | 4 (4) |
| Other | 11 (10) |
| Device or surgical procedure | 2 (2) |
| Referral and hospital stay | 7 (6) |
| **Diagnostic** | **34 (31)** |
| Imaging | 14 (13) |
| Laboratory | 12 (11) |
| Screening | 5 (5) |
| Other | 3 (3) |
| **Both diagnostic and therapeutic interventions** | **5 (5)** |

*Percentages do not add up to 100% in these categories because categories are not mutually exclusive.

**Factors**

In total, 404 unique factors were identified across the 111 included articles. Figure 2 shows the numbers of factors of the different levels; 158 factors (39%) on the individual provider; 82 factors (20%) on the individual patient level; 82 factors (20%) on the organizational context; 48 factors (12%) on the social context; and 34 factors (8%) on the economic and political context.

Of the 404 factors, 225 were classified as barriers (56%), 70 as facilitators (17%), 38 as both barrier and facilitator (9%) and 71 as determinant (18%). We first present the barriers, facilitators, and the factors that could be both a barrier and a facilitator (n=333) in more detail below and in Table 2. Thereafter, we describe determinants separately.

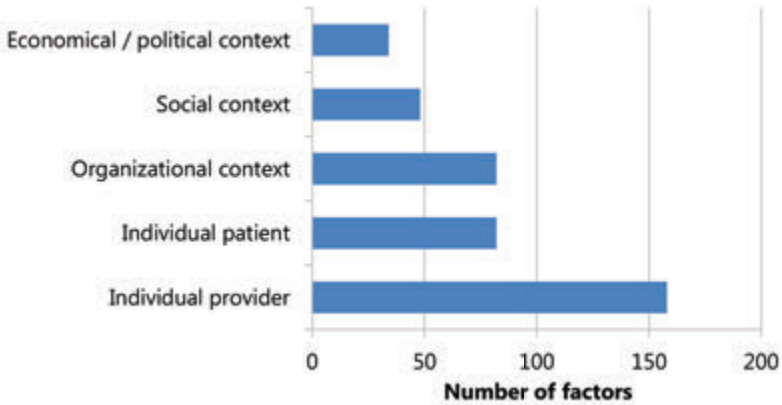



**Figure 2. Proportion of factors (N=404)**

**Table 2. Barriers and facilitators sub classification (n=333)**

| | Total n | % | Barriers n | % | Facilitators n | % | Both barrier and facilitator n | % |
|---|---|---|---|---|---|---|---|---|
| **Total provider characteristics** | **131** | **39.3** | **97** | **43.1** | **22** | **31.4** | **12** | **31.6** |
| Provider attitude | **72** | 21.6 | 53 | 23.6 | 11 | 15.7 | 8 | 21.1 |
| Provider knowledge | **43** | 12.9 | 34 | 15.1 | 7 | 10.0 | 2 | 5.3 |
| Provider behaviour | **11** | 3.3 | 6 | 2.7 | 4 | 5.7 | 1 | 2.6 |
| Individual provider characteristics | **5** | 1.5 | 4 | 1.8 | 0 | 0 | 1 | 2.6 |
| **Total patient characteristics** | **58** | **17.4** | **42** | **18.7** | **5** | **7.1** | **11** | **28.9** |
| Patient knowledge | **33** | 9.9 | 26 | 11.6 | 3 | 4.3 | 4 | 10.5 |
| Patient attitude | **13** | 3.9 | 8 | 3.6 | 2 | 2.9 | 3 | 7.9 |
| Patient behaviour | **7** | 2.1 | 4 | 1.8 | 0 | 0 | 3 | 7.9 |
| Individual patient characteristics | **5** | 1.5 | 4 | 1.8 | 0 | 0 | 1 | 2.6 |
| **Total social context characteristics** | **46** | **13.8** | **19** | **8.4** | **22** | **31.4** | **5** | **13.2** |
| Social context professional teams | **23** | 6.9 | 5 | 2.2 | 14 | 20.0 | 4 | 10.5 |
| Social context professional development | **12** | 3.6 | 7 | 3.1 | 5 | 7.1 | 0 | 0 |
| Social context professional networks | **11** | 3.3 | 7 | 3.1 | 3 | 4.3 | 1 | 2.6 |
| **Total organizational context characteristics** | **67** | **20.1** | **45** | **20.0** | **17** | **24.3** | **5** | **13.2** |
| Organizational context available resources | **28** | 8.4 | 22 | 9.8 | 4 | 5.7 | 2 | 5.3 |
| Organizational context structures and work routines | **24** | 7.2 | 11 | 4.9 | 10 | 14.3 | 3 | 7.9 |
| Organizational context processes | **15** | 4.5 | 12 | 5.3 | 3 | 4.3 | 0 | 0 |
| **Total economic political context** | **31** | **9.3** | **22** | **9.8** | **4** | **5.7** | **5** | **13.2** |
| Economic political context financial incentives | **27** | 8.1 | 20 | 8.9 | 3 | 4.3 | 4 | 10.5 |
| Economic political context legal regulatory measures | **2** | 0.6 | 0 | 0 | 1 | 1.4 | 1 | 2.6 |
| Economic political context segment of target groups | **2** | 0.6 | 2 | 0.9 | 0 | 0 | 0 | 0 |
| **Total** | **333** | **100** | **225** | **100** | **70** | **100** | **38** | **100** |

Framework according to Grol and Wensing[20,21]

**Individual provider (n=131; 39%)**

In terms of factors related to the individual provider, the most often identified factors were related to the attitude of the provider (n=72; 55%), followed by knowledge and skills (n=43; 33%).

Identified factors related to attitude included beliefs and opinions of healthcare providers, fear of medical errors, defensive attitude, motivation and commitment to restrict unnecessary care, and awareness of an agreement with guidelines. Among attitudes, the desire to meet expectations of the patients plays a major role. Facilitators to positive attitudes towards change that were named are a sense of ownership and participation in the project, a desire to restrict unnecessary care, and public commitment to change.

Articles also reported that attitudes can be influenced by a fear of medical error, litigation, public censure, and criticism from peers or supervisors. Other articles noted a more overall general defensive attitude towards medicine. For example, a study on reducing the use of antibiotics concluded that 'When there is uncertainty in any potentially infectious condition physicians tend to be cautious and prescribe an antibiotic if it could be at all beneficial.'[25]

Even if a provider has the necessary knowledge and attitudes for stopping with low-value care, behavior may still be difficult to change.[26] A few articles noted healthcare provider behavior as a factor, which is related to routines and habits. As with any type of behavioral modification, routines and habits in clinical practice can be difficult to change. Additionally, practical constraints, such as their workload and lack of time, play a role in a provider's ability to change their behavior.

Closely related to knowledge are experience and skills, which can be influenced by prior education and training. The most commonly reported skill was the provider-patient communication. Lack of communication skills needed to convince the patient that a test or treatment is not necessary and may be harmful, can pose a barrier. For example, while healthcare providers may have the knowledge that it is better to withhold from antibiotics for symptomatic relief of respiratory tract infections in children, changing their prescribing behavior may be difficult if they lack specific consulting skills to reassure patients without a prescription.[27]

**Individual patient (n=58; 17%)**

Factors related to patient knowledge were the most frequently reported patient related factors (n=33; 56%), followed by patient attitude (n=13; 22%). For all subcategories of factors related to the individual patient, the majority of the factors were identified as barriers (n=42; 72%), and a few facilitators (n=5; 9%). 19% of factors were identified as both a barrier and facilitator.

Patient knowledge, including patient expectation, was reported in the majority of the articles as a barrier (n=26; 45%), indicating that a lack of knowledge of the patient can pose a serious barrier to de-implementation. In addition to the role of the provider in giving adequate information on treatment options, some articles noted that patient knowledge can be influenced through media, internet, and advertisement from drug or medical device companies.[28,29]

In terms of patient attitude, some papers showed that patients express a preference for defensive medicine, perhaps stemming from anxiety, a false perception that they are at high risk, fear of complications of not intervening, or desire for diagnostic certainty and perceived control. One study noted that when offered a choice, many patients opt for more aggressive care than needed.[30] It was also identified that patient attitude can be influenced by prior experiences with the care practice. For example, reduction in symptoms after starting medication (whether it was related to the medication or not) may lead one to believe in the efficacy of medication.[31]

**Social context (n=46, 14%)**

In terms of social context, the majority of factors were related to professional teams (n=23; 50%); followed by professional development (n=12; 26%) and professional networks (n=11; 24%). Medical leadership was the most frequently recorded social context factor in the success of de-implementation.[32-36] These articles suggested that individuals who take an active role in quality improvement projects can positively influence the attitude of the team towards de-implementation, creating a positive culture where there is collaboration and good communication. A team approach is important to de-implementation as clinicians reported to be influenced by the expectation or requests from colleagues or to have been influenced by the knowledge, opinion, and action of their peers. Agreement on the appropriateness of interventions and the availability of clear guidelines at the level of medical associations can foster success of reducing agreed upon low-value care. Finally, healthcare providers may be influenced by pharmaceutical and medical device companies who have vested interests in seeing that their product is used.

**Organizational context (n=67; 20%)**

Available resources appeared to be the most important factor in the organizational context (n=28; 41%), followed by organizational structures and work routines (n=24; 36%) and organizational processes (n=15; 22%). Mainly barriers were identified in the organizational context (n=45; 67%), followed by facilitators (n=17; 25%) and factors that could be both barriers and facilitators (n=5; 7%).

Time was the most commonly reported resource factor, mainly as a barrier. Lack of time was often mentioned in reference to short consultation times, which pose a challenge to the in-depth provider-patient communication required for shared decision making. Another factor was the availability of resources. The ease of access to or simply the availability of interventions can influence their use. For example, the simple act of removing a checkbox for a specific blood test from a form results in less requests.[37,38]

In terms of organizational processes, several articles concluded that hospital or clinical practice databases play a key role in supporting quality improvement. The technical constraints of the database and the ease at which databases could be combined might either hinder or facilitate the ability to build in reminders into the system or monitor the quality of care and progress of de-implementation. Similarly, the right information needs to be available in the database (e.g. current prescriptions) to support the decision to withhold low-value care. Additionally, de-implementation was more difficult when it requires a change to the existing workflow or referral patters. Already existing automatic processes, such as the scheduling of (unnecessary) follow-up appointments or referrals can pose barriers.

**Economic and political context (n=31; 9%)**

Under the category of economic and political context, financial incentives were the largest group (n=27; 87%) followed by legal and regulatory measures and segments of the target group (both n=2). The latter included barriers related to involving diverse stakeholders and dealing with conflicting interests. The majority of the factors were barriers (n=22; 71%), whereas 4 factors (13%) were facilitators and 5 factors (16%) were both barriers and facilitators.

Financial incentives were found to be significant factors in the success of de-implementation. Financial incentives directed at the care provider were often mentioned, such as payment models which reward volume of care rather than those which hold them accountable for unnecessary care. Financial incentives directed at the patients were also mentioned, such as high co-payments and extensive insurance coverage leading patients to expect the providers to do something, such as run a diagnostic test, prescribe a medication, or referring them instead of sending them home. Factors related to the legal regulatory measures included barriers because of for example governmental reimbursement policies.

**Determinants**

Of the 71 determinants, 27 (38%) were categorized in provider factors and 24 determinants (34%) were related to patient characteristics. The provider characteristics

included mainly the age or clinical experience of the healthcare provider. Organizational factors (n=14; 20%) included demographic characteristics of hospitals or availability of staff, and resources. Social context factors and economic and political factors included both 3 determinants.

## Discussion

### Key findings

This evidence synthesis fills the knowledge gap on barriers and facilitators related to de-implementation or reducing of low-value care. In the 111 studies included in this review, over 400 factors are identified, spread over different subcategories. In addition to healthcare provider factors, many other factors are identified related to the patient, social context, organizational context and economical/political context.

Almost 40% of the factors identified were related to the individual healthcare provider and those were mainly related to attitude. This suggests that a de-implementation strategy based on provider education (focusing on knowledge) alone may be insufficient in many situations. Patient-provider communication and the desire to meet expectations of the patients play a major role. When faced with an uncertain outcome, clinicians prefer to avoid a greater unlikely loss than to incur a certain, but lesser, cost.[39]



The social, organizational, economic, and political context in which de-implementation takes place can also influence its success. Behavioral change is easier in a supportive environment; medical leadership and supervision on the de-implementation as well as positive constructive attitudes of the team towards de-implementation were facilitators.[40] Time was also a factor often mentioned; it may take longer to convince a patient that it is better to refrain from action than to request or prescribe low-value care. Focused patient information might help the healthcare provider in the consultation room.[41,42] Also of relevance to reducing low-value care is the problem of supplier-induced demand; financial incentives may encourage (or at least not dissuade) the provider to continue providing unnecessary treatment.[43]

For clinical practice, it is relevant to analyze the differences between factors influencing de-implementation and those influencing implementation. We compared our review to other reviews on barriers and facilitators influencing the practice of evidence based medicine[44-46] and a review on drivers of overdiagnosis.[47] It seems that patient-provider interaction, the fear of consequences of withholding a test or treatment, and financial incentives are more important factors in de-implementation than in implementation, although future research should investigate this more specifically.

Reviews by Cochrane et al. and Fischer et al. that focused primarily on implementation did not mention any patient-provider related factors.[44,45] Many articles in our review mentioned that patient preference, expectation, or request in combination with the physicians' communication skills and the time constraints of the consult were major barriers. Due to cognitive dissonance, physicians and patients alike may find it difficult to accept that a care practice which they believed to be effective is actually not. De-implementation may require longer and more difficult conversations with the patient.

The study on drivers of overdiagnosis noted that fears of uncertainty, ageing, death, and disease contribute to a culture of excess in medicine.[47] Our review found several references to fear, both at the patient level (defensive attitude), and the provider level (e.g. fear of consequences for patients' health, medical error, litigation). Emotional or extreme cases tend to stay in the memory and cause us to misjudge the actual frequency and magnitude of events.[12] Therefore, it is reasonable to infer that fear is a more prominent barrier to removal of excess (e.g. de-implementation) than implementation of a new test or treatment. This implies that stronger evidence is needed to convince healthcare providers that there is no harm in stopping with a certain care practice.

In terms of financial incentives, all three above mentioned reviews did not find as much evidence of financial incentives as playing a role in evidence-based medicine as we found in our review on de-implementation. This argues that supplier-induced demand in healthcare poses a major challenge to the reduction of low-value care.[43]

**Strengths and limitations**

A strength of this evidence synthesis was the broad search that resulted in a high number of included articles. Articles on de-implementation are difficult to find due to lack of consistent terminology; 43 different terms have been identified for de-implementation[9,10] and de-implementation articles are often described as articles on implementation of guidelines in which the guideline is to stop the low-value care. We believe that despite the possibility of missing relevant articles, a high degree of knowledge saturation has been reached.

An important limitation of this review is the exclusion of articles on disinvestment in which the motivation to stop was primarily financial. As a consequence, some macro-level factors, such as financial incentives may be underrepresented. Another limitation of our study could be the choice to use a predefined classification for barriers and facilitators to categorize qualitative data instead of a bottom up approach in which a new framework was developed based on the data.[48] On the other hand, using such

classification designed for implementation provided us insights into what might be specific to de-implementation. We also included articles that measured the effect of a de-implementation strategy, in which determinants related to the effect of the intervention were measured (classified as 'determinants' in our study). Such subgroup analysis or multivariable models might only include variables that are easy to measure (such as age and gender). This may result in an overrepresentation of specific variables, as you can only analyse the factors you have measured, whereas in other research designs, such as interviews or surveys, a broader range of factors were inventoried. For this reason, and due to the fact that those determinants are more difficult to categorize in barriers and facilitators, we described them separately.

**Implications for practice**

Once a service has been identified as low-value care, a first step towards reducing it should be to identify reasons why it (still) exists and to identify potential challenges to changing the current situation. The results of this study might help in identifying barriers and facilitators which would stimulate the development of a targeted strategy. In this overview, we used a narrow definition of barriers and facilitators with the intent to focus only on factors that could be targeted in de-implementations strategies. Several additional elements can influence the success of de-implementation, such as characteristics of the de-implementation strategy itself, the strength of the evidence against a clinical practice,[49] whether low-value care is only inefficient or if it also has negative health consequences,[50] or the type of change (e.g. removal, reduction, or replacement). Identifying factors that affect the influence of the effect of the de-implementation or the continuation of low-value care should be identified for each specific practice. This can be done through several methods including searching the literature, evaluating quantitative data on practice variation, and surveying or interviewing different stakeholders involved. Thereafter, a tailored strategy can be developed which takes into consideration who (patient or healthcare provider) or what level of organization (individual, context, or system) to target, and how behavioral change will be encouraged.



## Conclusions

This evidence synthesis provides insight into the range of factors affecting the success of de-implementation strategies. As most articles report factors on different levels, we conclude that multi-level de-implementation strategies might be necessary for effective reduction of low-value care. There is no one-size fits all solution: situation specific knowledge across all levels is important necessary for tailor-made de-implementation strategies.

## References


1. Morgan DJ, Brownlee S, Leppin AL, Kressin N, Dhruva SS, et al. Setting a research agenda for medical overuse. *BMJ.* 2015;351:h4534.
2. Colla CH, Mainor AJ, Hargreaves C, Sequist T, Morden N. Interventions Aimed at Reducing Use of Low-Value Health Services: A Systematic Review. *Med Care Res Rev.* 2017;74(5):507-550.
3. de Vries EF, Struijs JN, Heijink R, Hendrikx RJ, Baan CA. Are low-value care measures up to the task? A systematic review of the literature. *BMC Health Serv Res.* 2016;16(1):405.
4. Garner S, Littlejohns P. Disinvestment from low value clinical interventions: NICEly done? *BMJ.* 2011;343:d4519.
5. Levinson W, Kallewaard M, Bhatia RS, Wolfson D, Shortt S, et al. 'Choosing Wisely': a growing international campaign. *BMJ Qual Saf.* 2015;24(2):167-174.
6. Wammes JJ, van den Akker-van Marle ME, Verkerk EW, van Dulmen SA, Westert GP, et al. Identifying and prioritizing lower value services from Dutch specialist guidelines and a comparison with the UK do-not-do list. *BMC Med.* 2016;14(1):196.
7. Rosenberg A, Agiro A, Gottlieb M, Barron J, Brady P, et al. Early Trends Among Seven Recommendations From the Choosing Wisely Campaign. *JAMA Intern Med.* 2015;175:1913-1920.
8. Kost A, Genao I, Lee JW, Smith SR. Clinical decisions made in primary care clinics before and after choosing wiselyTM. *J Am Board Fam Med* 2015;28:471-474.
9. Niven DJ, Mrklas KJ, Holodinsky JK, Straus SE, Hemmelgarn BR, et al. Towards understanding the de-adoption of low-value clinical practices: a scoping review. *BMC Med.* 2015;13:255.
10. Gnjidic D, Elshaug AG. De-adoption and its 43 related terms: harmonizing low-value care terminology. *BMC Med.* 2015;13.
11. Roman BR, Asch DA. Faded promises: the challenge of deadopting low-value care. *Ann Intern Med* 2014;161(2):149-150.
12. Ubel PA, Asch DA. Creating value in health by understanding and overcoming resistance to de-innovation. *Health Aff (Millwood).* 2015;34(2):239-244.
13. Prasad V, Ioannidis JP. Evidence-based de-implementation for contradicted, unproven, and aspiring healthcare practices. *Implement Sci.* 2014;9:1.
14. Scott IA, Elshaug AG. Foregoing low-value care: how much evidence is needed to change beliefs? *Intern Med J.* 2013;43(2):107-109.
15. van Bodegom-Vos L, Davidoff F, Marang-van de Mheen PJ. Implementation and de-implementation: two sides of the same coin? *BMJ Qual Saf.* 2016.
16. Haas M, Hall J, Viney R, Gallego G. Breaking up is hard to do: why disinvestment in medical technology is harder than investment. *Aust Health Rev.* 2012;36(2):148-152.

17. Colla CH, Mainor AJ, Hargreaves C, Sequist T, Morden N. Interventions Aimed at Reducing Use of Low-Value Health Services: A Systematic Review. *Med Care Res Rev.* 2016.
18. Dixon-Woods M. Using framework-based synthesis for conducting reviews of qualitative studies. *BMC Med.* 2011;9:39.
19. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. *BMC Med Res Methodol.* 2009;9:59.
20. Grol R, Wensing M, Eccles MP, Davis D. *Improving patient care: the implementation of change in health care*. John Wiley & Sons; 2013.
21. Grol Richard PTM, Wensing M. *Implementation. Effective improvement of patient care*. Maarssen: Elsevier Gezondheidszorg; 2006.
22. Yates TD, Davis ME, Taylor YJ, Davidson L, Connor CD, et al. Not a magic pill: a qualitative exploration of provider perspectives on antibiotic prescribing in the outpatient setting. *BMC Fam Pract.* 2018;19(1):96.
23. Ralston SL, Atwood EC, Garber MD, Holmes AV. What Works to Reduce Unnecessary Care for Bronchiolitis? A Qualitative Analysis of a National Collaborative. *Acad Pediatr.* 2017;17(2):198-204.
24. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med Res Methodol.* 2012;12:181.
25. Arnold SR, Straus SE. Interventions to improve antibiotic prescribing practices in ambulatory care. *Cochrane Database Syst Rev.* 2005:CD003539.
26. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA*. 1999;282(15):1458-1465.
27. Andrews T, Thompson M, Buckley DI, Heneghan C, Deyo R, et al. Interventions to influence consulting and antibiotic use for acute respiratory tract infections in children: A systematic review and Meta-Analysis. *PLoS One.* 2012;7.
28. Melnick ER, Shafer K, Rodulfo N, Shi J, Hess EP, et al. Understanding Overuse of Computed Tomography for Minor Head Injury in the Emergency Department: A Triangulated Qualitative Study. *Acad Emerg Med.* 2015;22:1474-1483.
29. Pittenger K, Williams BL, Mecklenburg RS, Blackmore CC. Improving acute respiratory infection care through nurse phone care and academic detailing of physicians. *J Am Board Fam Med.* 2015;28:195-204.
30. Kanzaria HK, Brook RH, Probst MA, Harris D, Berry SH, Hoffman JR. Emergency physician perceptions of shared decision-making. *Acad Emerg Med*. 2015;22:399-405.
31. Vicens C, Bejarano F, Sempere E, Mateu C, Fiol F, et al. Comparative efficacy of two interventions to discontinue long-term benzodiazepine use: cluster randomised controlled trial in primary care. *Br J Psychiatry.* 2014;204:471-479.

32. Calderon-Margalit R, Mor-Yosef S, Mayer M, Adler B, Shapira SC. An administrative intervention to improve the utilization of laboratory tests within a university hospital. *International journal for quality in health care*. 2005;17(3):243-248.
33. King R, Michelman M, Curran V, Bean J, Rowden P, Lindsey J. Patient-centered approach to ensuring appropriateness of care through blood management. *South Med J*. 2013;106:362-368.
34. Miyakis S, Karamanof G, Liontos M, Mountokalakis Theodore D. Factors contributing to inappropriate ordering of tests in an academic medical department and the effect of an educational feedback strategy. *Postgrad Med J*. 2006;82(974):823-829.
35. Williams I, Harlock J, Robert G, Mannion R, Brearley S, Hall K. Decommissioning health care: identifying best practice through primary and secondary research – a prospective mixed-methods study. *NIHR Journals Library Health Services and Delivery Research*. 2017;7:07.
36. Graham ID, Logan J, Davies B, Nimrod C. Changing the use of electronic fetal monitoring and labor support: a case study of barriers and facilitators. *Birth*. 2004;31:293-301.
37. Bailey J, Jennings A, Parapia L. Change of pathology request forms can reduce unwanted requests and tests. *J Clin Pathol* 2005;58(8):853-855.
38. Shalev V, Chodick G, Heymann AD. Format change of a laboratory test order form affects physician behavior. *Int J Med Inform*. 2009;78(10):639-644.
39. Simianu VV, Grounds MA, Joslyn SL, LeClerc JE, Ehlers AP, et al. Understanding clinical and non-clinical decisions under uncertainty: a scenario-based survey. *BMC Med Inform Decis Mak*. 2016;16(1):153.
40. Chauhan BF, Jeyaraman M, Mann AS, Lys J, Skidmore B, et al. Behavior change interventions and policies influencing primary healthcare professionals' practice-an overview of reviews. *Implement Sci*. 2017;12(1):3.
41. Meeker D, Knight TK, Friedberg MW, Linder JA, Goldstein NJ, et al. Nudging guideline-concordant antibiotic prescribing: A randomized clinical trial. *JAMA Intern Med*. 2014;174:425-431.
42. Krol N, Wensing M, Haaijer-Ruskamp F, Muris JW, Numans ME, et al. Patient-directed strategy to reduce prescribing for patients with dyspepsia in general practice: a randomized trial. *Aliment Pharmacol Ther*. 2004;19:917-922.
43. Hensher M, Tisdell J, Zimitat C. "Too much medicine": Insights and explanations from economic theory and research. *Soc Sci Med*. 2017;176:77-84.
44. Cochrane LJ, Olson CA, Murray S, Dupuis M, Tooman T, Hayes S. Gaps between knowing and doing: understanding and assessing the barriers to optimal health care. *J Contin Educ Health Prof*. 2007;27(2):94-102.
45. Fischer F, Lange K, Klose K, Greiner W, Kraemer A. Barriers and Strategies in Guideline Implementation-A Scoping Review. *Healthcare (Basel)*. 2016;4(3).

46. Zwolsman S, te Pas E, Hooft L, Wieringa-de Waard M, van Dijk N. Barriers to GPs' use of evidence-based medicine: a systematic review. *Br J Gen Pract*. 2012;62(600):e511-521.
47. Pathirana T, Clark J, Moynihan R. Mapping the drivers of overdiagnosis to potential solutions. *BMJ*. 2017;358:j3879.
48. Espeland A, Baerheim A. Factors affecting general practitioners' decisions about plain radiography for back pain: implications for classification of guideline barriers--a qualitative study. *BMC Health Serv Res*. 2003;3(1):8.
49. Howard DH, Gross CP. Producing Evidence to Reduce Low-Value Care. *JAMA Intern Med*. 2015;175(12):1893-1894.
50. Bentley TG, Effros RM, Palar K, Keeler EB. Waste in the U.S. Health care system: a conceptual framework. *Milbank Q*. 2008;86(4):629-659.

## Supplemental appendix

### Search strategy

Date: September 12, 2018

*Databases*

**Medline**

((((obsole* or (("not" or "no longer") adj (effective or essential or efficient)) or ineffective or uneffective or low-value or overuse* or inappropriate or "old habits") adj4 ("health system" or healthcare or care or policy or policies or practice or technology or procedure* or treatment* or intervention* or "health services" or strateg* or "clinical use" or referral* or diagnosis or regulatory or approach or prescrib* or therap*)) or (overtest* or overdiagnos*)) adj5 (reduce or avoid or minimise or discontinu* or minimise or decreas* or stop or stopping or revers* or replace* or avert or "trim down" or (cut adj (down or back)) or substitute or decrement)).mp. OR ("de-implementation" or deimplementation or "do-not-do" or "deadopt*" or decommission*).mp.

**Embase**

#1 ((((obsole* or (("not" or "no longer") adj (effective or essential or efficient)) or ineffective or uneffective or low-value or overuse* or inappropriate or "old habits") adj4 ("health system" or healthcare or care or policy or policies or practice or technology or procedure* or treatment* or intervention* or "health services" or strateg* or "clinical use" or referral* or diagnosis or regulatory or approach or prescrib* or therap*)) or (overtest* or overdiagnos*)) adj5 (reduce or avoid or minimise or discontinu* or minimise or decreas* or stop or stopping or revers* or replace* or avert or "trim down" or (cut adj (down or back)) or substitute or decrement)).mp. OR ("de-implementation" or deimplementation or "do-not-do" or "deadopt*" or decommission*).mp.
#2 limit (conference abstract OR conference paper OR conference review)

**Rx for change**

"de-adoption" OR "Decrease use" OR "abandoning" OR "Discontinue" OR "discontinuation" OR "Abandon" OR "Reassess" OR "reassessment" OR "Obsolete" OR "Medical reversal" OR "Re-invest" OR "Withdraw*" OR "de-implementation" OR "Reduc*" OR "Decline in use" OR "Health technology reassessment" OR "Change in use" OR "De-implement*" OR "De-list" OR "De-commission" OR "Do not do" OR "Reallocation" OR "relinquishing" OR "Over use" OR "Stop" OR "Inappropriate use" OR "Relinquish*" OR "Ineffective" OR "Misuse" OR "Re-appraisal" OR "Re-prioritization" OR "Clinical redesign" OR "Disadoption" OR "Redeploy" OR "Reversal" OR "Drop in use" OR "stopping"

*Websites*

- Canadian Agency for Drugs and Technologies in Health - https://www.cadth.ca/resources/rx-for-change/database/browse
- Agency for Healthcare Research and Quality - http://www.ahrq.gov/
- Right Care - http://www.rightcare.nhs.uk/
- National Institute for Health and Care Excellence - https://www.nice.org.uk/
- Choosing Wisely - http://www.choosingwisely.org/
- Kingsfund - http://www.kingsfund.org.uk/
- Nuffield Trust - http://www.nuffieldtrust.org.uk/

## Detailed information of included studies

| Study | Study design | Goal article |
|---|---|---|
| Anton, 2007[1] | Before-after | Measure effectiveness deimplementation |
| Aspinal, 2007[2] | Cohort | Measure effectiveness deimplementation |
| Awad, 2006[3] | RCT | Measure effectiveness deimplementation |
| Bailey, 2005[4] | ITS | Measure effectiveness deimplementation |
| Banerjee, 2011[5] | Survey | Identification factors influencing low-value care |
| Barnes, 2017[6] | Other, mixed methods | Identification factors influencing low-value care |
| Batuwitage, 2006[7] | Before-after | Measure effectiveness deimplementation |
| Bauchner, 1999[8] | Survey | Identification factors influencing low-value care |
| Belongia, 2001[9] | Before-after | Measure effectiveness deimplementation |
| Bishop, 2017[10] | Qualitative | Identification factors influencing low-value care |
| Brady, 2017[11] | Non-RCT | identification factors influencing low-value care |
| Calderon-Margalit, 2005[12] | Before-after | Measure effectiveness deimplementation |
| Chirima, 2016[13] | Cohort | Measure effectiveness deimplementation |
| Clyne, 2017[14] | Qualitative | identification factors influencing low-value care |
| Colla, 2017[15] | Other, mixed methods | Identification factors influencing low-value care |
| Cossette, 2016[16] | Cohort | Measure effectiveness deimplementation |
| Davies, 2002[17] | RCT | Measure effectiveness deimplementation |
| De Miguel, 2000[18] | ITS | Measure effectiveness deimplementation |
| Dempsey, 2014[19] | Qualitative | Identification factors influencing low-value care |
| Dhalla, 2002[20] | Before-after | Measure effectiveness deimplementation |
| Duane, 2016[21] | Qualitative | Identification factors influencing low-value care |
| Duffy, 1992[22] | Cohort | Identification factors influencing low-value care |
| Fagan, 2014[23] | Before-after | Measure effectiveness deimplementation |
| Flottorp, 2003[24] | Other, mixed methods | identification factors influencing low-value care |
| Freeborn, 1997[25] | Before-after | Measure effectiveness deimplementation |
| Gershengorn, 2013[26] | Cohort | identification factors influencing low-value care |
| Gjelstad, 2013[27] | RCT | Measure effectiveness deimplementation |
| Gordon, 2000[28] | ITS | Measure effectiveness deimplementation |
| Graham, 2004[29] | Qualitative | Identification factors influencing low-value care |
| Green, 2018[30] | Qualitative | Identification factors influencing low-value care |
| Gupta, 2013[31] | Before-after | Measure effectiveness deimplementation |
| Hammond, 2009[32] | qualitative | identification factors influencing low-value care |
| Hamzat, 2012[33] | ITS | Measure effectiveness deimplementation |
| Harris, 2003[34] | NonRCT | Measure effectiveness deimplementation |
| Hatam, 2010[35] | Chart review | identification factors influencing low-value care |
| Hooper, 2009[36] | Before-after | Measure effectiveness deimplementation |

| | Factors | | | | |
|---|---|---|---|---|---|
| **Intervention** | **Provider** | **Patient** | **Social context** | **Organisational** | **Economic-political** |
| Diagnostic resource use (referrals) | X | | | X | |
| Antibiotic | X | X | X | X | |
| Antibiotic | X | X | X | X | |
| Diagnostic laboratory | | | | X | |
| Therapeutic other | X | | | | |
| Diagnostic laboratory | X | | | X | |
| Gastric acid suppresives | X | | | | |
| Antibiotic | X | X | X | X | |
| Antibiotic | | X | | | |
| Several therapeutic and diagnostic interventions | X | X | X | X | |
| Diagnostic laboratory | | | X | X | |
| Diagnostic laboratory antibiotic | | X | X | | |
| Benzodiazepines | X | | | | |
| Several therapeutic and diagnostic interventions | X | | | X | |
| Polypharmacy | X | | | | |
| Diagnostic imaging | X | | X | | |
| Therapeutic other | X | | | | |
| Antibiotic | | X | X | X | |
| Polypharmacy | X | X | | X | |
| Antibiotic | X | X | | X | X |
| Therapeutic other | X | X | | X | |
| Antibiotic | | X | | | |
| Several interventions | X | | X | X | |
| Diagnostic imaging | X | X | | | |
| Therapeutic other | | X | | X | |
| Antibiotic | X | X | | | |
| Benzodiazepines | X | X | | X | |
| Diagnostic imaging | X | X | | X | |
| Several therapeutic and diagnostic interventions | | X | X | | X |
| Gastric acid suppresives | | X | | | |
| Diagnostic resource use (referrals) | | X | X | X | |
| Gastric acid suppresives | | X | | | |
| Antibiotic | X | X | | | |
| Therapeutic other | X | | | X | X |
| Benzodiazepines | | | | | X |

| | | |
|---|---|---|
| Hussein, 2010[37] | Survey | Identification factors influencing low-value care |
| Hutchinson, 1999[38] | Cohort | identification factors influencing low-value care |
| Juzych, 2005[39] | NonRCT | Measure effectiveness deimplementation |
| Kakkar, 2004[40] | Before-after | Measure effectiveness deimplementation |
| Kanzaria, 2015[41] | Other, mixed methods | identification factors influencing low-value care |
| Kaul, 2015[42] | Survey | identification factors influencing low-value care |
| King, 2013[43] | Before-after | identification factors influencing low-value care |
| Klein, 2017[44] | Survey | identification factors influencing low-value care |
| Kline, 2017[45] | Survey | identification factors influencing low-value care |
| Kruse, 2015[46] | Cohort | Measure effectiveness deimplementation |
| Kulawik, 2009[47] | Before-after | Measure effectiveness deimplementation |
| Kumar, 2003[48] | Qualitative | Identification factors influencing low-value care |
| Lambert-Kerzner, 2018[49] | Qualitative | Identification factors influencing low-value care |
| Lee, 2017[50] | Survey | Identification factors influencing low-value care |
| Liao, 2017[51] | Survey | Identification factors influencing low-value care |
| Lin, 2016[52] | Qualitative | Measure effectiveness deimplementation |
| Lin, 2017[53] | Survey | identification factors influencing low-value care |
| Linder, 2003[54] | Chart review | identification factors influencing low-value care |
| Linder, 2007[55] | Before-after | Measure effectiveness deimplementation |
| Linsky, 2015[56] | Qualitative | Identification factors influencing low-value care |
| Lipitz-Snyderman, 2016[57] | Cohort | Measure effectiveness deimplementation |
| Liu, 2012[58] | ITS | Measure effectiveness deimplementation |
| Lum, 2017[59] | Qualitative | Identification factors influencing low-value care |
| Macfarlane, 1997[60] | Survey | Identification factors influencing low-value care |
| Mahalingam, 2015[61] | Before-after | Measure effectiveness deimplementation |
| Mainous Iii, 1998[62] | Cohort | Identification factors influencing low-value care |
| Maughan, 2015[63] | Qualitative | Identification factors influencing low-value care |
| May, 2006[64] | Before-after | Measure effectiveness deimplementation |
| McKay, 2017[65] | Qualitative | Identification factors influencing low-value care |
| McNicholl, 2017[66] | Chart review | Measure effectiveness deimplementation |
| Melnick, 2015[67] | Qualitative | Identification factors influencing low-value care |
| Miyakis, 2006[68] | Before-after | Both identifying factors influencing de-implementation or the continuation of low-value care AND measure the effectiveness of de-implementation |
| Murray, 2000[69] | Survey | Identification factors influencing low-value care |
| Murthy, 2006[70] | Survey | Identification factors influencing low-value care |
| Nachnani, 2009[71] | Chart review | Identification factors influencing low-value care |
| Palmer, 1997[72] | Survey | Identification factors influencing low-value care |
| Perz, 2002[73] | Before-after | Measure effectiveness deimplementation |
| Pittenger, 2015[74] | ITS | Measure effectiveness deimplementation |
| Pollock, 2000[75] | Qualitative | Identification factors influencing low-value care |
| Raghunath, 2005[76] | Qualitative | Identification factors influencing low-value care |
| Ralston, 2013[77] | ITS | Measure effectiveness deimplementation |

| | | | | | |
|---|---|---|---|---|---|
| Gastric acid suppresives | X | | | | |
| Antibiotic | X | X | | | X |
| Antibiotic | X | | | | |
| Blood transfusion | | | | | X |
| Diagnostic laboratory | | X | | | |
| diagnostic resource use (referrals) | X | | | X | X |
| Blood transfusion | | | X | X | |
| Antibiotic | X | X | | X | |
| Diagnostic imaging | X | X | | | |
| Diagnostic screening | X | | | | |
| Therapeutic other | | | X | | |
| Antibiotic | X | | X | | |
| Device surgical procedure | X | X | | | |
| Antibiotic | X | | | X | X |
| Several interventions | | | | | X |
| Diagnostic imaging | X | X | X | X | X |
| Antibiotic | X | | | | |
| Antibiotic | | | | X | |
| Antibiotic | | | | X | |
| Polypharmacy | X | X | | X | X |
| Several interventions | X | | | | |
| Antibiotic | X | | | | |
| Antibiotic | X | X | X | | |
| Antibiotic | X | X | | | |
| Diagnostic resource use (referrals) | X | | | | |
| Antibiotic | X | | | | |
| Several interventions | X | | | X | |
| Diagnostic laboratory | | | X | X | |
| Therapeutic other | X | | X | X | |
| Several interventions | | X | | | |
| Diagnostic imaging | X | X | | X | |
| Diagnostic laboratory | X | X | | X | |

| | | | | | |
|---|---|---|---|---|---|
| Antibiotic | | X | | | |
| Gastric acid suppresives | X | | | | |
| Gastric acid suppresives | X | X | | | |
| Antibiotic | | X | | | |
| Antibiotic | | | | X | |
| Antibiotic | | | | X | |
| Gastric acid suppresives | | X | | | |
| Gastric acid suppresives | X | X | | X | |
| Diagnostic resource use (referrals) | | | X | | |

| | | |
|---|---|---|
| Ralston, 2016[78] | Before-after | Measure effectiveness deimplementation |
| Ralston, 2017[79] | Other, mixed methods | Identification factors influencing low-value care |
| Reed - Antibiotics in acute bronchitis, 2015 [80] | Other, mixed methods | Identification factors influencing low-value care |
| Reed - Carotid artery stenosis, 2015[81] | Other, mixed methods | Identification factors influencing low-value care |
| Reed - low back pain, 2015[82] | Other, mixed methods | Identification factors influencing low-value care |
| Reed - Pap testing, 2015[83] | Other, mixed methods | Identification factors influencing low-value care |
| Reed - Percutaneous coronary interventions, 2015[84] | Other, mixed methods | Identification factors influencing low-value care |
| Reed - Preoperative stress testing, 2015[85] | Other, mixed methods | Identification factors influencing low-value care |
| Reed – Headache, 2015[86] | Other, mixed methods | Identification factors influencing low-value care |
| Rosenthal, 2018[87] | Cohort | Identification factors influencing low-value care |
| Samore, 2005[88] | RCT | Identification factors influencing low-value care |
| Sawan, 2016[89] | Qualitative | Identification factors influencing low-value care |
| Schmidt, 2018[90] | Cohort | Identification factors influencing low-value care |
| Seager, 2005[91] | RCT | Measure effectiveness deimplementation |
| Sedrak, 2016[92] | Survey | Identification factors influencing low-value care |
| Shepperd, 2013[93] | Qualitative | Identification factors influencing low-value care |
| Sloane, 2014[94] | Before-after | Measure effectiveness deimplementation |
| Soria-Aledo, 2012[95] | Before-after | Measure effectiveness deimplementation |
| Steinke, 2000[96] | Cohort | Measure effectiveness deimplementation |
| Stinnett-Donnelly, 2015[97] | Other, mixed methods | Measure effectiveness deimplementation |
| Thomas, 2002[98] | Before-after | Measure effectiveness deimplementation |
| Tierny, 1990[99] | RCT | Measure effectiveness deimplementation |
| Urfer, 2016[100] | Before-after | Measure effectiveness deimplementation |
| van Bodegom-Vos, 2016[101] | Qualitative | Identification factors influencing low-value care |
| Vegting, 2012[102] | Before-after | Measure effectiveness deimplementation |
| Voorn, 2017[103] | RCT | Measure effectiveness deimplementation |
| Walker, 2001[104] | Qualitative | Identification factors influencing low-value care |
| Weddle, 2017[105] | Before-after | Measure effectiveness deimplementation |
| Wermeling, 2014[106] | Qualitative | Identification factors influencing low-value care |
| Williams, 2017[107] | Survey | Measure effectiveness deimplementation |
| Winchester, 2014[108] | Chart review | Identification factors influencing low-value care |
| Winchester, 2017[109] | Before-after | Measure effectiveness deimplementation |
| Yates, 2018[110] | Qualitative | Identification factors influencing low-value care |
| Zabarsky, 2008[111] | Before-after | Measure effectiveness deimplementation |

RCT: randomized controlled trial; ITS: interrupted times series

| | | | | | |
|---|---|---|---|---|---|
| Diagnostic resource use (referrals) | | | X | | |
| Therapeutic other | X | | X | X | X |
| Antibiotic | | X | | X | X |
| Diagnostic screening | | X | | X | X |
| Diagnostic imaging | X | X | | X | X |
| Diagnostic screening | X | X | X | X | X |
| Device surgical procedure | X | X | | | X |
| Diagnostic other | X | X | | X | X |
| Diagnostic imaging | X | X | | X | X |
| Several interventions | | | | X | |
| Antibiotic | X | X | | | |
| Benzodiazepines | X | | X | | |
| Antibiotic | X | X | | | |
| antibiotic | | X | | | |
| Diagnostic laboratory | X | | | X | |
| Several therapeutic and diagnostic interventions | X | X | | X | X |
| Antibiotic | X | | X | | |
| Diagnostic resource use (referrals) | X | X | | X | |
| Antibiotic | | X | | X | |
| Several interventions | | | | X | |
| Antibiotic | | | | X | |
| Diagnostic laboratory | | | | | X |
| Polypharmacy | | X | | | |
| Blood transfusion | X | | | | |
| Diagnostic laboratory | X | | | | |
| Blood transfusion | X | | | X | |
| Antibiotic | X | | | | |
| Antibiotic | | X | | | |
| Gastric acid suppresives | X | | | | |
| Several therapeutic and diagnostic interventions | | | X | | |
| Diagnostic imaging | | X | | | |
| Diagnostic imaging | X | | X | | |
| Antibiotic | | X | | X | |
| Antibiotic | X | | X | X | |

**References**


1. Anton P, Peiro S, Aranaz JM, Calpena R, Compan A, et al. Effectiveness of a physician-oriented feedback intervention on inappropriate hospital stays. *J Epidemiol Community Health.* 2007;61:128-134.
2. Aspinall SL, Metlay JP, Maselli JH, Gonzales R. Impact of hospital formularies on fluoroquinolone prescribing in emergency departments. *Am J Manag Care.* 2007;13(5):241-248.
3. Awad AI, Eltayeb IB, Baraka OZ. Changing antibiotics prescribing practices in health centers of Khartoum State, Sudan. *Eur J Clin Pharmacol.* 2006;62(2):135-142.
4. Bailey J, Jennings A, Parapia L. Change of pathology request forms can reduce unwanted requests and tests. *J Clin Pathol.* 2005;58(8):853-855.
5. Banerjee SC, Siriwardena AN, Iqbal M. What influences pre-hospital cannulation intentions in paramedics? An application of the theory of reasoned action. *J Eval Clin Pract.* 2011;17:84-90.
6. Barnes GD, Misirliyan S, Kaatz S, Jackson EA, Haymart B, et al. Barriers and facilitators to reducing frequent laboratory testing for patients who are stable on warfarin: a mixed methods study of de-implementation in five anticoagulation clinics. *Implement Sci.* 2017;12(1):87.
7. Batuwitage BT, Kingham JGC, Morgan NE, Bartlett RL. Inappropriate prescribing of proton pump inhibitors in primary care. *Postgrad Med J.* 2007;83:66-68.
8. Bauchner H, Pelton SI, Klein JO. Parents, physicians, and antibiotic use. *Pediatrics.* 1999;103(2):395-401.
9. Belongia EA, Sullivan BJ, Chyou PH, Madagame E, Reed KD, Schwartz B. A community intervention trial to promote judicious antibiotic use and reduce penicillin-resistant Streptococcus pneumoniae carriage in children. *Pediatrics.* 2001;108(3):575-583.
10. Bishop TF, Cea M, Miranda Y, Kim R, Lash-Dardia M, et al. Academic physicians' views on low-value services and the choosing wisely campaign: A qualitative study. *Healthcare.* 2017;5(1-2):17-22.
11. Brady H, Piggott L, Dunne SS, O'Connell NH, Dunne CP. Clustered interventions to reduce inappropriate duplicate laboratory tests in an Irish tertiary hospital. *Clin Biochem.* 2017:19.
12. Calderon-Margalit R, Mor-Yosef S, Mayer M, Adler B, Shapira SC. An administrative intervention to improve the utilization of laboratory tests within a university hospital. *International journal for quality in health care.* 2005;17(3):243-248.
13. Chirima T, Esterhuizen T. Inappropriate antibiotic prescribing: A cross-sectional study in Swaziland. *SA Pharmaceutical Journal.* 2016;83(2):44-49.

14. Clyne B, Cooper JA, Boland F, Hughes CM, Fahey T, Smith SM. Beliefs about prescribed medication among older patients with polypharmacy: A mixed methods study in primary care. *Br J Gen Pract.* 2017;67(660):e507-e518.
15. Colla CH, Mainor AJ. Choosing Wisely Campaign: Valuable For Providers Who Knew About It, But Awareness Remained Constant, 2014-17. *Health Aff.* 2017;36(11):2005-2011.
16. Cossette B, Bergeron J, Ricard G, Ethier JF, Joly-Mischlich T, et al. Knowledge Translation Strategy to Reduce the Use of Potentially Inappropriate Medications in Hospitalized Elderly Adults. *J Am Geriatr Soc.* 2016;64(12):2487-2494.
17. Davies B, Hodnett E, Hannah M, O'Brien-Pallas L, Pringle D, et al. Fetal health surveillance: a community-wide approach versus a tailored intervention for the implementation of clinical practice guidelines. *CMAJ.* 2002;167(5):469-474.
18. De M, Vargas E, Portoles A, Puerro M, Terleira AI, Moreno A. Modification of albumin use pattern after an educational intervention. *Dis Manag Health Out.* 2000;8:43-50.
19. Dempsey PP, Businger AC, Whaley LE, Gagne JJ, Linder JA. Primary care clinicians' perceptions about antibiotic prescribing for acute bronchitis: a qualitative study. *BMC Fam Pract.* 2014;15:194.
20. Dhalla IA, Anderson GM, Mamdani MM, Bronskill SE, Sykora K, Rochon PA. Inappropriate prescribing before and after nursing home admission. *J Am Geriatr Soc.* 2002;50:995-1000.
21. Duane S, Beatty P, Murphy AW, Vellinga A. Exploring Experiences of Delayed Prescribing and Symptomatic Treatment for Urinary Tract Infections among General Practitioners and Patients in Ambulatory Care: A Qualitative Study. *Antibiotics*. 2016;5(3):15.
22. Duffy SQ, Farley DE. The protracted demise of medical technology. The case of intermittent positive pressure breathing. *Med Care.* 1992;30:718-736.
23. Fagan M, Lindbaek M, Reiso H, Berild D. A simple intervention to reduce inappropriate ciprofloxacin prescribing in the emergency department. *Scand J Infect Dis.* 2014;46:481-485.
24. Flottorp S, Havelsrud K, Oxman AD. Process evaluation of a cluster randomized trial of tailored interventions to implement guidelines in primary care--why is it so hard to change practice? *Fam pract.* 2003;20(3):333-339.
25. Freeborn DK, Shye D, Mullooly JP, Eraker S, Romeo J. Primary care physicians' use of lumbar spine imaging tests: effects of guidelines and practice pattern feedback. *J Gen Intern Med.* 1997;12(10):619-625.
26. Gershengorn HB, Wunsch H. Understanding changes in established practice: pulmonary artery catheter use in critically ill patients. *Crit Care Med.* 2013;41:2667-2676.

27. Gjelstad S, Høye S, Straand J, Brekke M, Dalen I, Lindbæk M. Improving antibiotic prescribing in acute respiratory tract infections: cluster randomised trial from Norwegian general practice (prescription peer academic detailing (Rx-PAD) study). *BMJ.* 2013;347:f4403-f4403.
28. Gordon DB, Jones HD, Goshman LM, Foley DK, Bland SE. A quality improvement approach to reducing use of meperidine. *Jt Comm J Qual Improv.* 2000;26(12):686-699.
29. Graham ID, Logan J, Davies B, Nimrod C. Changing the use of electronic fetal monitoring and labor support: a case study of barriers and facilitators. *Birth.* 2004;31:293-301.
30. Green AR, Tung M, Segal JB. Older Adults' Perceptions of the Causes and Consequences of Healthcare Overuse: A Qualitative Study. *J Gen Intern Med.* 2018;33(6):892-897.
31. Gupta R, Marshall J, Munoz JC, Kottoor R, Jamal MM, Vega KJ. Decreased acid suppression therapy overuse after education and medication reconciliation. *Int J Clin Pract.* 2013;67:60-65.
32. Hammond CL, Pinnington LL, Phillips MF. A qualitative examination of inappropriate hospital admissions and lengths of stay. *BMC Health Serv Res.* 2009;9:44.
33. Hamzat H, Sun H, Ford JC, MacLeod J, Soiza RL, Mangoni AA. Inappropriate prescribing of proton pump inhibitors in older patients: Effects of an educational strategy. *Drugs Aging.* 2012;29:681-690.
34. Harris Robert H, MacKenzie Thomas D, Leeman-Castillo B, Corbett Kitty K, Batal Holly A, et al. Optimizing antibiotic prescribing for acute respiratory tract infections in an urban urgent care clinic. *J Gen Intern Med.* 2003;18(5):326-334.
35. Hatam N, Askarian M, Sarikhani Y, Ghaem H. Necessity of admissions in selected teaching university affiliated and private hospitals during 2007 in Shiraz, Iran. *Arch Iran Med.* 2010;13(3):230-234.
36. Hooper S, Bruno R, Sharpe M, Tahmindjis A. Alprazolam prescribing in Tasmania: A two-fold intervention designed to reduce inappropriate prescribing and concomitant opiate prescription. *Australas Psychiatry.* 2009;17:300-305.
37. Hussain S, Stefan M, Visintainer P, Rothberg M. Why do physicians prescribe stress ulcer prophylaxis to general medicine patients? *South Med J.* 2010;103:1103-1110.
38. Hutchinson JM, Foley RN. Method of physician remuneration and rates of antibiotic prescription. *CMAJ.* 1999;160:1013-1017.
39. Juzych Nadia S, Banerjee M, Essenmacher L, Lerner Stephen A. Improvements in antimicrobial prescribing for treatment of upper respiratory tract infections through provider education. *J Gen Intern Med.* 2005;20(10):901-905.

40. Kakkar N, Kaur R, Dhanoa J. Improvement in fresh frozen plasma transfusion practice: results of an outcome audit. *Transfus Med.* 2004;14:231-235.
41. Kanzaria HK, Brook RH, Probst MA, Harris D, Berry SH, Hoffman JR. Emergency physician perceptions of shared decision-making. *Acad Emerg Med.* 2015;22:399-405.
42. Kaul S, Kirchhoff AC, Morden NE, Vogeli CS, Campbell EG. Physician response to patient request for unnecessary care. *Am J Manag Care.* 2015;21:823-832.
43. King R, Michelman M, Curran V, Bean J, Rowden P, Lindsey J. Patient-centered approach to ensuring appropriateness of care through blood management. *South Med J.* 2013;106:362-368.
44. Klein EY, Martinez EM, May L, Saheed M, Reyna V, Broniatowski DA. Categorical Risk Perception Drives Variability in Antibiotic Prescribing in the Emergency Department: A Mixed Methods Observational Study. *J Gen Intern Med.* 2017;32(10):1083-1089.
45. Kline KP, Shaw L, Beyth RJ, Plumb J, Nguyen L, et al. Perceptions of patients and providers on myocardial perfusion imaging for asymptomatic patients, choosing wisely, and professional liability. *BMC Health Serv Res.* 2017;17(1):553.
46. Kruse GR, Khan SM, Zaslavsky AM, Ayanian JZ, Sequist TD. Overuse of colonoscopy for colorectal cancer screening and surveillance. *J Gen Intern Med.* 2015;30:277-283.
47. Kulawik D, Sands JJ, Mayo K, Fenderson M, Hutchinson J, et al. Focused vascular access education to reduce the use of chronic tunneled hemodialysis catheters: results of a network quality improvement initiative. *Semin Dial.* 2009;22:692-697.
48. Kumar S, Little P, Britten N. Why do general practitioners prescribe antibiotics for sore throat? Grounded theory interview study. *BMJ.* 2003;326(7381):138-138.
49. Lambert-Kerzner A, Maynard C, McCreight M, Ladebue A, Williams KM, et al. Assessment of barriers and facilitators in the implementation of appropriate use criteria for elective percutaneous coronary interventions: a qualitative study. *BMC Cardiovasc Disord.* 2018;18(1):164.
50. Lee TH, Wong JG, Lye DC, Chen MI, Loh VW, et al. Medical and psychosocial factors associated with antibiotic prescribing in primary care: survey questionnaire and factor analysis. *Br J Gen Pract.* 2017;67(656):e168-e177.
51. Liao JM, Navathe AS, Schapira MM, Weissman A, Mitra N, Asch DA. Penalizing Physicians for Low-Value Care in Hospital Medicine: A Randomized Survey. *J Hosp Med.* 2017:22.
52. Lin IB, Coffin J, O'Sullivan PB. Using theory to improve low back pain care in Australian Aboriginal primary care: a mixed method single cohort pilot study. *BMC Fam Pract.* 2016;17:44.

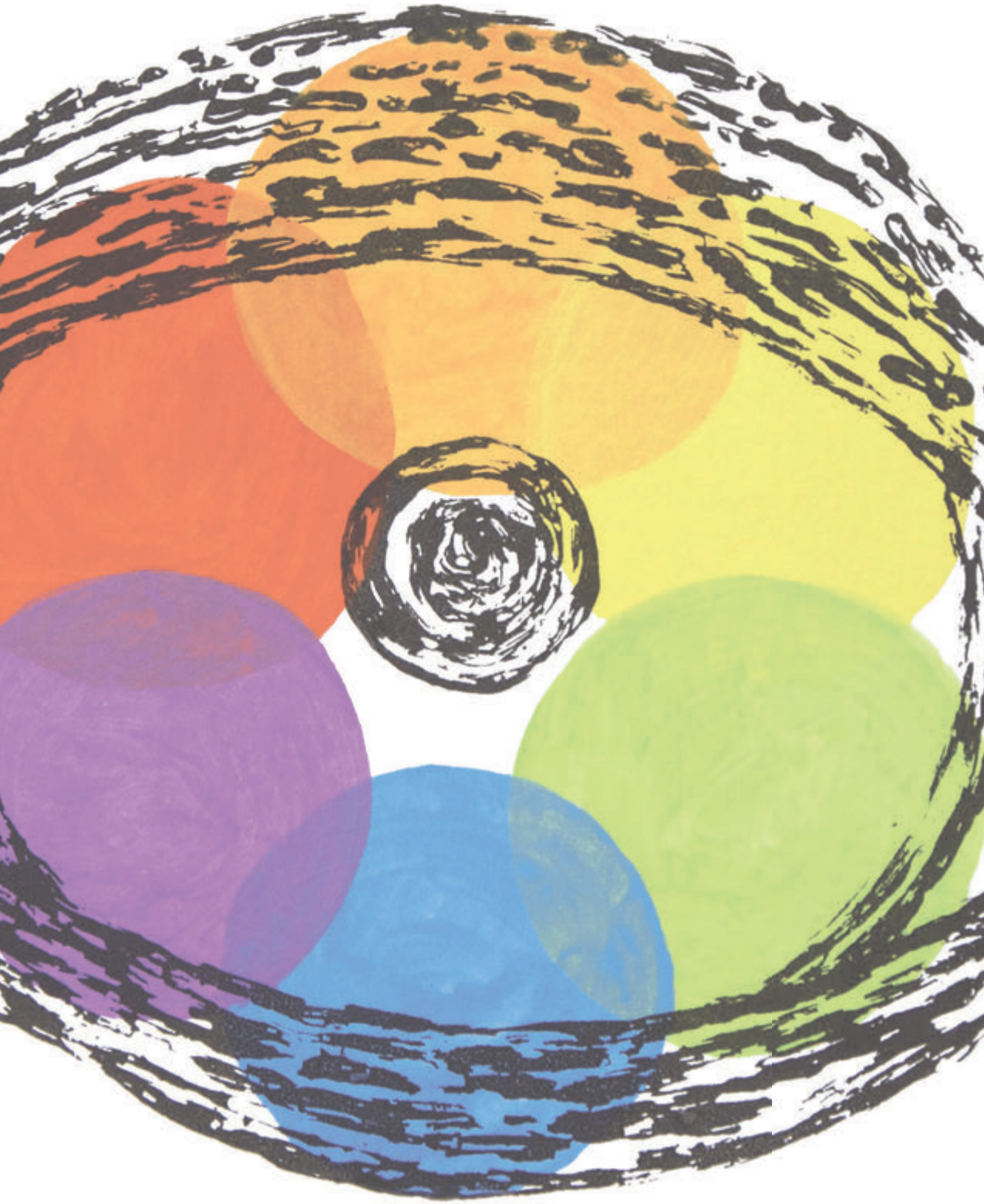
53. Lin MP, Nguyen T, Probst MA, Richardson LD, Schuur JD. Emergency Physician Knowledge, Attitudes and Behavior regarding ACEP's Choosing Wisely Recommendations: A Survey Study. *Acad Emerg Med.* 2017.
54. Linder JA, Singer DE, Stafford RS. Association between antibiotic prescribing and visit duration in adults with upper respiratory tract infections. *Clin Ther.* 2003;25:2419-2430.
55. Linder J, Schnipper JL, Volk LA, Tsurikova R, Palchuk M, et al. Clinical decision support to improve antibiotic prescribing for acute respiratory infections: results of a pilot study. *AMIA* 2007;Annual Symposium proceedings / AMIA Symposium. AMIA Symp:468-472.
56. Linsky A, Simon SR, Marcello TB, Bokhour B. Clinical provider perceptions of proactive medication discontinuation. *Am J Manag Care.* 2015;21:277-283.
57. Lipitz-Snyderman A, Sima CS, Atoria CL, Elkin EB, Anderson C, et al. Physician-Driven Variation in Nonrecommended Services Among Older Adults Diagnosed With Cancer. *JAMA Intern Med.* 2016;176(10):1541-1548.
58. Liu Z, Abdullah A, Baskin L, Lewis G, Kelter G, Naugler C. An intervention to reduce laboratory utilization of referred-out tests. *Lab Med.* 2012;43(5 S):164-167.
59. Lum EPM, Page K, Nissen L, Doust J, Graves N. Australian consumer perspectives, attitudes and behaviours on antibiotic use and antibiotic resistance: a qualitative study with implications for public health policy and practice. *BMC Public Health.* 2017;17(1):799.
60. Macfarlane J, Holmes W, Macfarlane R, Britten N. Influence of patients' expectations on antibiotic management of acute lower respiratory tract illness in general practice: questionnaire study. *BMJ.* 1997;315(7117):1211-1214.
61. Mahalingam S, Seymour N, Pepper C, Tostevin P, Oakeshott P. Reducing inappropriate referrals to secondary care: our experiences with the ENT Emergency clinic. *Qual Prim Care.* 2014;22:251-255.
62. Mainous Iii AG, Hueston WJ, Love MM, Evans ME, Finger R. An evaluation of statewide strategies to reduce antibiotic overuse. *Fam Med.* 2000;32:22-29.
63. Maughan BC, Baren JM, Merchant RM. Choosing wisely in emergency medicine: A national survey of academic leaders. *Acad Emerg Med.* 2015;1:S144.
64. May TA, Clancy M, Critchfield J, Ebeling F, Enriquez A, et al. Reducing unnecessary inpatient laboratory testing in a teaching hospital. *Am J Clin Pathol.* 2006;126(2):200-206.
65. McKay VR, Margaret Dolcini M, Hoffer LD. The dynamics of de-adoption: a case study of policy change, de-adoption, and replacement of an evidence-based HIV intervention. *Transl Behav Med.* 2017:10.
66. McNicholl IR, Gandhi M, Hare CB, Greene M, Pierluissi E. A Pharmacist-Led Program to Evaluate and Reduce Polypharmacy and Potentially Inappropriate Prescribing in Older HIV-Positive Patients. *Pharmacotherapy.* 2017:10.

67. Melnick ER, Shafer K, Rodulfo N, Shi J, Hess EP, et al. Understanding Overuse of Computed Tomography for Minor Head Injury in the Emergency Department: A Triangulated Qualitative Study. *Acad Emerg Med.* 2015;22:1474-1483.
68. Miyakis S, Karamanof G, Liontos M, Mountokalakis Theodore D. Factors contributing to inappropriate ordering of tests in an academic medical department and the effect of an educational feedback strategy. *Postgrad Med J* 2006;82(974):823-829.
69. Murray S, Del M, O'Rourke P. Predictors of an antibiotic prescription by GPs for respiratory tract infections: a pilot. *Fam Pract.* 2000;17(5):386-388.
70. Murthy SK, Kauldher S, Targownik LE. Physicians' approaches to the use of gastroprotective strategies in low-risk non-steroidal anti-inflammatory drug users. *Aliment Pharmacol Ther.* 2006;23:1365-1372.
71. Nachnani JS, Bulchandani D, Moormeier J, Foxworth J. Patient and physician predictors of inappropriate acid-suppressive therapy (AST) use in hospitalized patients. *J Hosp Med.* 2009;4:E10-E14.
72. Palmer DA, Bauchner H. Parents' and physicians' views on antibiotics. *Pediatrics.* 1997;99:E6.
73. Perz Joseph F, Craig Allen S, Coffey Christopher S, Jorgensen Daniel M, Mitchel E, et al. Changes in antibiotic prescribing for children after a community-wide campaign. *JAMA.* 2002;287(23):3103-3109.
74. Pittenger K, Williams BL, Mecklenburg RS, Blackmore CC. Improving acute respiratory infection care through nurse phone care and academic detailing of physicians. *J Am Board Fam Med.* 2015;28:195-204.
75. Pollock K, Grime J. Strategies for reducing the prescribing of proton pump inhibitors (PPIs): patient self-regulation of treatment may be an under-exploited resource. *Soc Sci Med.* 2000;51:1827-1839.
76. Raghunath AS, Hungin AP, Cornford CS, Featherstone V. Use of proton pump inhibitors: an exploration of the attitudes, knowledge and perceptions of general practitioners. *Digestion.* 2005;72:212-218.
77. Ralston S, Garber M, Narang S, Shen M, Pate B, et al. Decreasing unnecessary utilization in acute bronchiolitis care: results from the value in inpatient pediatrics network. *J Hosp Med.* 2013;8(1):25-30.
78. Ralston SL, Garber MD, Rice-Conboy E, Mussman GM, Shadman KA, et al. A Multicenter Collaborative to Reduce Unnecessary Care in Inpatient Bronchiolitis. *Pediatrics.* 2016;137:1-9.
79. Ralston SL, Atwood EC, Garber MD, Holmes AV. What Works to Reduce Unnecessary Care for Bronchiolitis? A Qualitative Analysis of a National Collaborative. *Acad Pediatr.* 2017;17(2):198-204.
80. Reed SJ, Pearson S. *Choosing Wisely® Recommendation Analysis: Prioritizing Opportunities for Reducing Inappropriate Care. Antibiotic use for acute bronchitis.* Institute for clinical and economic review (ICER); 2015/01// 2015.

81. Reed SJ, Pearson S. *Choosing Wisely® Recommendation Analysis: Prioritizing Opportunities for Reducing Inappropriate Care. Carotid artery stenosis screening in asymptomatic patients.* Institute for clinical and economic review (ICER); 2015/12// 2015.
82. Reed SJ, Pearson S. *Choosing Wisely® Recommendation Analysis: Prioritizing Opportunities for Reducing Inappropriate Care. Imaging for nonspecific low back pain.* Institute for clinical and economic review (ICER); 2015/12// 2015.
83. Reed SJ, Pearson S. *Choosing Wisely® Recommendation Analysis: Prioritizing Opportunities for Reducing Inappropriate Care. Pap testing.* Institute for clinical and economic review (ICER); 2015/01// 2015.
84. Reed SJ, Pearson S. *Choosing Wisely® Recommendation Analysis: Prioritizing Opportunities for Reducing Inappropriate Care. Percutaneous coronary intervention for stable ischemic heart disease.* Institute for clinical and economic review (ICER); 2015/12// 2015.
85. Reed SJ, Pearson S. *Choosing Wisely® Recommendation Analysis: Prioritizing Opportunities for Reducing Inappropriate Care. Preoperative stress testing.* Institute for clinical and economic review (ICER); 2015/12// 2015.
86. Reed SJ, Pearson S. *Choosing Wisely® Recommendation Analysis: Prioritizing Opportunities for Reducing Inappropriate Care. Imaging for uncomplicated headache.* Institute for clinical and economic review (ICER); 2015/12// 2015.
87. Rosenthal MB, Colla CH, Morden NE, Sequist TD, Mainor AJ, et al. Overuse and insurance plan type in a privately insured population. *Am J Manag Care.* 2018;24(3):140-146.
88. Samore MH, Bateman K, Alder SC, Hannah E, Donnelly S, et al. Clinical decision support and appropriateness of antimicrobial prescribing: A randomized trial. *JAMA.* 2005;294:2305-2314.
89. Sawan M, Jeon YH, Fois RA, Chen TF. Exploring the link between organizational climate and the use of psychotropic medicines in nursing homes: A qualitative study. *Res Social Adm Pharm.* 2016:05.
90. Schmidt ML, Spencer MD, Davidson LE. Patient, Provider, and Practice Characteristics Associated with Inappropriate Antimicrobial Prescribing in Ambulatory Practices. *Infect Control Hosp Epidemiol.* 2018;39(3):307-315.
91. Seager JM, Howell-Jones RS, Dunstan FD, Lewis MAO, Richmond S, Thomas DW. A randomised controlled trial of clinical outreach education to rationalise antibiotic prescribing for acute dental pain in the primary care setting. *Br Dent J.* 2005;201(4):217-222; discussion 216.
92. Sedrak MS, Patel MS, Ziemba JB, Murray D, Kim EJ, et al. Residents' self-report on why they order perceived unnecessary inpatient laboratory tests. *J Hosp Med.* 2016;11(12):869-872.
93. Shepperd S, Adams R, Hill A, Garner S, Dopson S. Challenges to using evidence from systematic reviews to stop ineffective practice: An interview study. *J Health Serv Res Policy.* 2013;18:160-166.

94. Sloane PD, Zimmerman S, Reed D, Beeber AS, Chisholm L, et al. Antibiotic prescribing in 4 assisted-living communities: Incidence and potential for improvement. *Infect Control Hosp Epidemiol.* 2014;35:S62-S68.
95. Soria-Aledo V, Carrillo-Alcaraz A, Flores-Pastor B, Moreno-Egea A, Carrasco-Prats M, Aguayo-Albasini JL. Reduction in inappropriate hospital use based on analysis of the causes. *BMC Health Serv Res.* 2012;12:361.
96. Steinke DT, Bain DJ, MacDonald TM, Davey PG. Practice factors that influence antibiotic prescribing in general practice in Tayside. *J Antimicrob Chemother.* 2000;46(3):509-512.
97. Stinnett-Donnelly JM, Stevens PG, Hood VL. Developing a high value care programme from the bottom up: a programme of faculty-resident improvement projects targeting harmful or unnecessary care. *BMJ Qual Saf.* 2015.
98. Thomas AR, Cieslak PR, Strausbaugh LJ, Fleming DW. Effectiveness of pharmacy policies designed to limit inappropriate vancomycin use: A population-based assessment. *Infect Control Hosp Epidemiol.* 2002;23:683-688.
99. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. *N Engl J Med.* 1990;322(21):1499-1504.
100. Urfer M, Elzi L, Dell-Kuster S, Bassetti S. Intervention to Improve Appropriate Prescribing and Reduce Polypharmacy in Elderly Patients Admitted to an Internal Medicine Unit. *PLoS One.* 2016;11(11):e0166359.
101. van Bodegom-Vos L, Davidoff F, Marang-van de Mheen PJ. Implementation and de-implementation: two sides of the same coin? *BMJ Qual Saf.* 2016:10.
102. Vegting IL, van Beneden M, Kramer MH, Thijs A, Kostense PJ, Nanayakkara PW. How to save costs by reducing unnecessary testing: lean thinking in clinical practice. *Eur J Intern Med.* 2012;23(1):70-75.
103. Voorn VMA, Marang-van de Mheen PJ, van der Hout A, Hofstede SN, So-Osman C, et al. The effectiveness of a de-implementation strategy to reduce low-value blood management techniques in primary hip and knee arthroplasty: a pragmatic cluster-randomized controlled trial. *Implement Sci.* 2017;12(1):72.
104. Walker AE, Grimshaw JM, Armstrong EM. Salient beliefs and intentions to prescribe antibiotics for patients with a sore throat. *Br J Health Psychol.* 2001;6:347-360.
105. Weddle G, Goldman J, Myers A, Newland J. Impact of an Educational Intervention to Improve Antibiotic Prescribing for Nurse Practitioners in a Pediatric Urgent Care Center. *J Pediatr Health Care.* 2017;31(2):184-188.
106. Wermeling M, Himmel W, Behrens G, Ahrens D. Why do GPs continue inappropriate hospital prescriptions of proton pump inhibitors? A qualitative study. *Eur J Gen Pract.* 2014;20:174-180.

107. Williams I, Harlock J, Robert G, Mannion R, Brearley S, Hall K. Decommissioning health care: identifying best practice through primary and secondary research – a prospective mixed-methods study. *NIHR Journals Library Health Services and Delivery Research.* 2017;7:07.
108. Winchester DE, Hymas J, Meral R, Nguyen D, Dusaj R, et al. Clinician-dependent variations in inappropriate use of myocardial perfusion imaging: Training, specialty, and location. *J Nucl Cardiol.* 2014;21:598-604.
109. Winchester DE, Schmalfuss C, Helfrich CD, Beyth RJ. A specialty-specific, multimodality educational quality improvement initiative to deimplement rarely appropriate myocardial perfusion imaging. *Open Heart.* 2017;4(1):e000589.
110. Yates TD, Davis ME, Taylor YJ, Davidson L, Connor CD, et al. Not a magic pill: a qualitative exploration of provider perspectives on antibiotic prescribing in the outpatient setting. *BMC Fam Pract.* 2018;19(1):96.
111. Zabarsky TF, Sethi AK, Donskey CJ. Sustained reduction in inappropriate treatment of asymptomatic bacteriuria in a long-term care facility through an educational intervention. *Am J Infect Control.* 2008;36:476-480.



# Chapter 7

## Effectiveness of strategies to reduce low-value care: a systematic review of de-implementation studies

Pauline Heus
Jan-Willem Weenink
Simone A van Dulmen
Christiana A Naaktgeboren
Eva W Verkerk
Maarten J van der Laan
Rudolf B Kool
Lotty Hooft

*Manuscript in preparation*

## Abstract

### Background

Low-value care is healthcare leading to no or little clinical benefit for the patient. It is unclear what the best (combinations of) interventions to reduce low-value care are. We aimed to quantify and compare the effectiveness of de-implementation strategies and identify which characteristics are related to the reduction of low-value care.

### Methods

Medline, Embase and Rx for Change databases were searched (1990 – September 2018) for randomised controlled trials evaluating a strategy to reduce low-value care. Additional publications were identified through searching websites of healthcare organizations and checking references. Risk of bias was assessed using the Cochrane Risk of Bias tool. We explored associations between strategy characteristics and effectiveness of de-implementation strategies.

### Results

Forty-nine randomised controlled trials were included, of which 38 (78%) addressed therapeutic low-value care practices. Compared to usual care, de-implementation led to a median relative reduction in the use of a low-value healthcare practice of 13% (IQR 9% - 27%). A smaller effect was found for strategies addressing a single target with a single intervention (n=5; 5%, IQR 5% – 20%). In reducing therapeutic low-value care, targeting the strategy to patients tends to achieve a larger effect (20% [IQR 10% - 43%]) compared to strategies were no patients were addressed (median 11% [IQR 7% - 21%]). Strategies containing audit and feedback had a larger median difference than strategies without this intervention (16% [IQR 9% - 27%] vs. 8% [IQR 6% - 13%]).

### Conclusions

A majority of de-implementation strategies achieved a considerable reduction of low-value care, especially those applying a multifaceted intervention. It seems worthwhile to consider audit and feedback and patient directed interventions. Details regarding sustainability of effect are often lacking, which is essential information needed for interpretation and application of findings.

## Introduction

Low-value care is healthcare that has no or little clinical benefit for the patient, considering the costs, the risks, available alternatives, and patient preferences.[1,2] Although hampered by the lack of clear definitions and international consensus, estimates of the volume of low-value care range from 10% to 30%.[2-4] Estimates up to 89% have been reported for specific healthcare practices.[5,6]

Low-value care and strategies to reduce it increasingly receive attention. In the last decade, several initiatives have been launched that list practices that doctors and patients should question or withhold.[7-10] Yet, raising awareness by presenting lists is not enough to reduce the use of these practices.[11,12] Previous research on changing behaviour showed that active rather than passive dissemination strategies are more likely to be effective.[13,14] With regard to reducing low-value care, however, it is unclear which active strategies are the best, as a first scoping review on this topic concluded.[15] This scoping review revealed a considerable body of literature. Apart from describing the terminology and frameworks used, the authors proposed a model to guide the process of reducing a low-value healthcare practice. In addition, they identified several knowledge gaps and one of their recommendations was to undertake a more detailed evidence synthesis to quantify the effectiveness of strategies applied to reduce low-value care.



A systematic review of active interventions aimed at reducing the use of low-value care (de-implementation) indicated that multicomponent interventions are potentially more effective than single-component interventions, especially when addressing both patients and clinicians.[2] This overview was descriptive, without comparing the absolute or relative measures of the effect of de-implementation strategies. Furthermore, observational studies without a parallel control group were also included in this review, making it hard to draw strong conclusions about effectiveness of strategies.

In this systematic review, our aim was to quantify and compare the effectiveness of de-implementation strategies across studies and to identify characteristics related to the reduction of low-value care, based on the best available evidence coming from randomised controlled trials. Our findings will contribute to the evidence-base needed for developing effective and sustainable de-implementation strategies to improve quality of care.

## Methods

### Data Sources and Searches

An information specialist searched EMBASE, MEDLINE, and Rx for Change databases on September 12th, 2018. Search terms included synonyms for de-implementation and low-value care. Websites of healthcare quality improvement organizations were also searched. Reference lists of articles screened on full text were used as an additional source of potentially relevant studies. Details regarding the search are available in Additional file 1.

### Study Selection

Randomised controlled trials (RCTs) in which a strategy aimed at reducing low-value care was studied, that were published after 1990 in English, German, French, or Dutch, were eligible. For protocols identified by the search, it was checked whether the study had been published as a full text. We included studies assessing the effect of a strategy on the incidence of low-value care, e.g. new prescriptions or test orders. Studies evaluating the cessation of long-term medication use (discontinuation in the context of an individual patient's care) were excluded. Studies on guideline adherence were only included when it was possible to extract information on reduction of healthcare practices. Pairs of authors independently screened titles and abstracts, and subsequently full texts of potentially eligible publications (CN, EV, JWW, PH, and SvD). Discrepancies were resolved through discussion and, when necessary, a third author was consulted.

### Data Extraction and Critical appraisal

Data of eligible publications were extracted by one author and checked by a second author (CN, EV, JWW, LH, MvdL, PH, RK, and SvD). To ensure consistency between the reviewers, we developed and piloted a standardised electronic data extraction form that included study characteristics (low-value care being de-implemented, targets and components of the de-implementation strategy, study design) and effect measures (outcomes). We distinguished four different target levels of a de-implementation strategy (hereafter just called targets): provider, patient, organizational context (including social context) and healthcare system, based on the categorization by Grol et al.[16] The components used in de-implementation strategies were classified in nine categories based on the taxonomy provided by the Cochrane Effective Practice and Organisation of Care (EPOC) Group (see Additional file 2).[17,18]

Two authors independently assessed risk of bias by applying the Cochrane Risk of bias tool (CN, EV, JWW, LH, MvdL, PH, SvD or RK).[19] In addition to the seven domains of this

tool, three issues specific to cluster randomised designs were assessed: recruitment bias, unit of analysis error, and concern regarding baseline imbalances.[20-23]

**Analyses**

To quantify and compare the effect of de-implementation strategies across studies (with both dichotomous and continuous outcomes), we calculated the relative changes in use of the low-value healthcare practice for each study arm. This requires either a reported relative change between baseline and post-intervention (i.e. after applying the de-implementation strategy) or data to calculate this (i.e. volume of low-value care measured pre- and post-intervention). If actual low-value care (care that was provided inappropriately) was not measured, total volume of care was used instead. Effectiveness was determined by taking the difference in relative changes between study arms (de-implementation strategy vs. usual care or other de-implementation strategy). Medians with interquartile ranges were used to summarize the effectiveness of strategies across studies.

When studies compared more than one de-implementation strategy to usual care, we included the data of the most complex strategy defined by the most interventions and/or targets. When a study evaluated more than two low-value care practices (e.g., various laboratory tests), the low-value care practice with the median relative reduction was taken. In case of two low-value care practices we selected the one with the largest relative reduction.

Differences in the effect of de-implementation strategies (i.e. relative reduction [with IQR]) were explored for several subgroups: type of low-value care (either diagnostic or therapeutic); number of targeted groups and intervention categories; whether the strategies were tailor-made based on pre-identified barriers and facilitators; type of outcome measured (total volume of care or actual low-value care); and overall risk of bias (on a study-level). Studies were classified as low risk of bias when they had 1) an adequate random sequence generation, 2) scored a low risk of bias for all three domains related to cluster randomised designs, and 3) no high risk of bias due to unconcealed allocation, detection bias, attrition bias, or reporting bias, with unclear risk of bias for a maximum of two of these domains. Furthermore, to evaluate relative effectiveness of strategies we selected studies that directly compared de-implementation strategies.

Besides the above mentioned analysis of the effect of de-implementation post-intervention at short term, we also assessed the available data on sustainability of effects.

Analyses were performed in R (version 3.6.0)[24] and Review Manager software[25] was used for generating the risk of bias figures.

## Results

### Search results

The search identified 4111 records to screen for eligibility (Figure 1). Based on title and abstract 3306 records were excluded. Full text assessment resulted in exclusion of an additional 756 records. Main reasons for exclusion were not evaluating a de-implementation strategy (n=424), or not being a randomised controlled trial (n=137). In total, 49 studies were included. An overview of included studies and their characteristics is provided in Additional file 3, table S1.

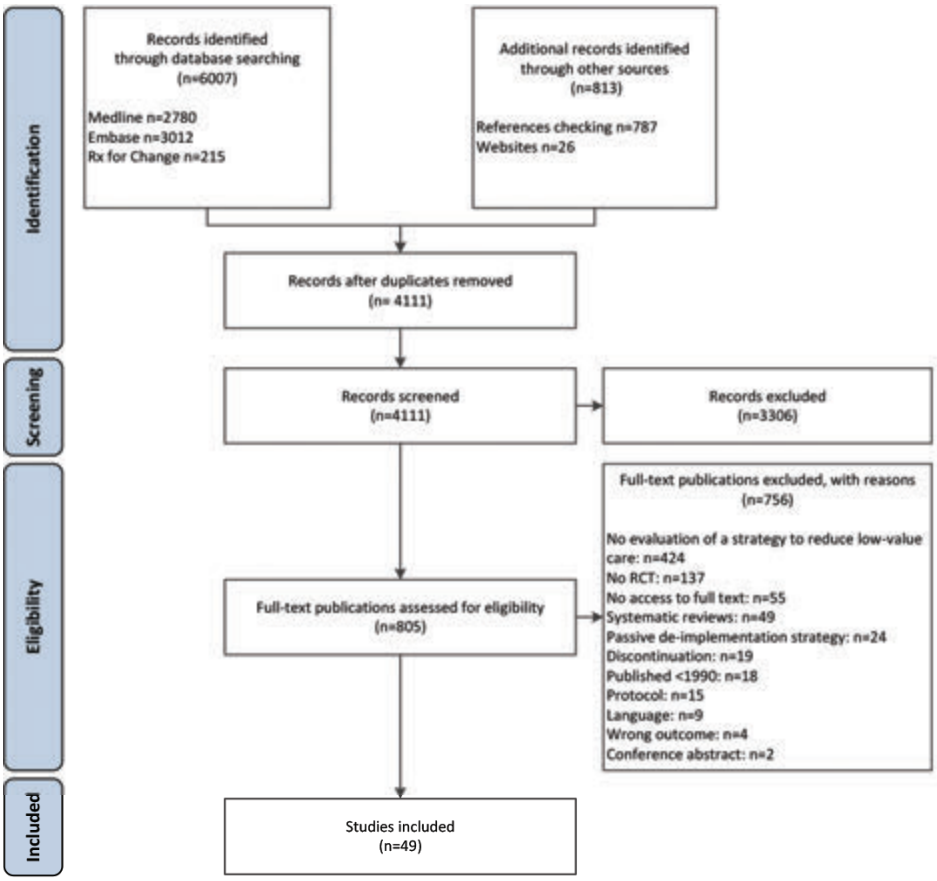



**Figure 1. Flow chart of selection process**

**Characteristics of included studies**

All but four (8%) of the included studies had a cluster-randomised design with randomisation on the level of healthcare providers (n=15; 31%), healthcare centres or groups of healthcare providers (n=27; 55%), or communities (n=3; 6%). Thirty-three (67%) were multicentre studies, 10 (20%) were community studies, and 6 (12%) took place in a single centre. Most studies were conducted in North America (n=27; 55%) or Europe (n=18; 37%). Strategies were compared to usual care in 43 studies (88%) and to another active de-implementation strategy in 11 studies (22%). The total adds up to more than n=49, as six studies fell into both categories.

*Types of low-value care being de-implemented*

The majority of the studies (n=38; 78%) addressed therapeutic low-value care (n=35 medication; n=3 non-medication) and were performed in a primary care or outpatient setting (n=36; 74%) (Table 1). The definition of low-value care was mainly based on guidelines from regional or national institutes (n=30; 61%) or evidence from literature (n=6; 12%), however nine studies (18%) were not explicit about this. In 42 (86%) studies the main goal was to reduce or not routinely provide the low-value care practice, rather than to abandon it completely.

*De-implementation strategies*

De-implementation strategies were classified according to two key variables: intervention(s) and targeted audience (Table 2). Nine strategies (18%) addressed a single target with a single intervention. Six of these strategies targeted healthcare providers, of which four consisted of reminders (including decision support tools). Additionally, 27 strategies (55%) addressed a single target as well (healthcare providers in all), but used a combination of interventions (multifaceted). More than half of them (n=15) combined education (meetings and/or distribution of materials) with audit and feedback (see Additional file 3, table S2). Another 13 multifaceted strategies (27%) were directed at multiple targets, of which eight at both healthcare providers and patients. A combination of provider education, audit and feedback, and a patient directed educational interventions was most often used (n=4). Multifaceted strategies addressed a median number of 2 (IQR 2 to 3) intervention categories.

**Table 1. Details regarding low-value care**
**(See Additional file 3, table S1 for characteristics of individual studies)**

| | All studies n=49 | Diagnostic healthcare practices n=11 (22%) | | Therapeutic healthcare practices n=38 (78%) | |
|---|---|---|---|---|---|
| | | *Imaging* | *4 (36)* | Medication | 35 (92) |
| | | *Laboratory tests* | *3 (27)* | *Antibiotics* | *27 (71)* |
| | | *Test ordering* | *2 (18)* | *Medication in general* | *5 (13)* |
| | | *Pathology* | *1 (9)* | *Benzodiazepines* | *2 (5)* |
| | | *Screening* | *1 (9)* | *Antidepressants* | *1 (3)* |
| | | | | Non-medication | 3 (8) |
| | | | | *Hospital utilization* | *1 (3)* |
| | | | | *Blood transfusion* | *1 (3)* |
| | | | | *Fetal monitoring* | *1 (3)* |
| **Setting** | | | | | |
| Primary care, outpatient services | 36 (74) | 9 (82) | | 27 (71) | |
| Hospital | 7 (14) | 1 (9) | | 6 (16) | |
| *Academic* | *2* | *1* | | *1* | |
| Long term care facility | 3 (6) | 0 | | 3 (8) | |
| Other, mixed | 3 (6) | 1 (9) | | 2 (5) | |
| **Definition of low-value based on** | | | | | |
| Guidelines | 30 (61) | 6 (55) | | 24 (63) | |
| Literature (reference provided) | 6 (12) | 1 (9) | | 5 (13) | |
| Panel | 4 (8) | 1 (9) | | 3 (8) | |
| Not specified | 9 (18) | 3 (27) | | 6 (16) | |
| **Aim** | | | | | |
| Reduce / Provide not routinely | 42 (86) | 10 (91) | | 32 (84) | |
| Stop | 4 (8) | 1 (9) | | 3 (8) | |
| Combination | 3 (6) | 0 | | 3 (8) | |

n (%)

**Table 2. Details of the evaluated de-implementation strategies with regard to interventions and targets**

| | Single target, single intervention | Single target, combination of interventions | Multiple targets**, combination of interventions | All strategies*** |
|---|---|---|---|---|
| | (N=9) | (N=27) | (N=13) | (N=49) |
| ***Intervention categories* and targets*** | | | | |
| **Targeted at provider** | **6/9 (67)** | **27/27 (100)** | **12/13 (92)** | **45/49 (92)** |
| - educational meetings (e.g. lectures, workshops, conferences) | 1/6 (17) | 16/27 (59) | 11/12 (92) | 28/45 (62) |
| - distribution of educational material (e.g. publications, guidelines, pocket cards) | NA | 23/27 (85) | 9/12 (75) | 32/45 (71) |
| - reminders (including decision support tools) | 4/6 (67) | 9/27 (33) | 3/12 (25) | 16/45 (36) |
| - audit and feedback | 1/6 (17) | 19/27 (70) | 7/12 (58) | 27/45 (60) |
| - financial interventions | NA | NA | NA | |
| **Targeted at patient** | **2/9 (25)** | **NA** | **12/13 (92)** | **14/49 (30)** |
| **Targeted at organisational context** | **1/9 (13)** | **NA** | **4/13 (31)** | **5/49 (10)** |
| - organisational interventions (redefining roles, multidisciplinary teams, appliances, test ordering procedures and forms) | 1/1 (100) | NA | 4/4 (100) | 5/5 (100) |
| - structural interventions (changing setting of care, e.g. from hospital to general practice) | NA | NA | NA | NA |
| **Targeted at healthcare system** | **NA** | **NA** | **1/13 (8)** | **1/49 (2)** |
| - regulatory interventions | NA | NA | NA | |
| - financial interventions | NA | NA | 1/1 (100) | 1/1 (100) |

n/N(%)

NA: not available (no studies)

* Based on taxonomy provided by the Cochrane Effective Practice and Organisation of Care (EPOC) Group

** Provider and patient n=8 (62%); provider and organisational context n=1 (8%); patient and organisational context n=1 (8%); provider, patient and organisational context n=2 (15%); provider, patient and healthcare system n=1 (8%)

***As a strategy can have more than one target, numbers add up to more than 49

*Outcome assessment*
All studies assessed the effect of de-implementation strategies on use of a healthcare practice considered to be of low-value. They measured either the effect on total volume of care (n=36; 74%), and/or the effect on the volume of actual low-value care (n=18; 37%). For assessing the effects, most studies used clinical data (e.g., electronic health records) (n=20; 41%), administrative data or registries (n=17; 35%), or both (n=8; 16%). In the remaining studies, a survey was used (n=3; 6%) and in one study the source of outcome data was unclear.

**Risk of bias**
Within most bias domains, the majority of studies were judged to be at a low risk of bias (see Additional file 3, figure S1). However, as interventions could not be blinded, a high risk of performance bias was considered for 43 (88%) of the studies. Details on randomisation procedure were not provided in 23 (47%) (sequence generation) and 32 (65%) (allocation concealment) randomised controlled trials, leading to an unclear risk of selection bias. Risk of reporting bias was judged to be unclear in 35 (71%) trials, as study protocols were not available, and high in 3 (6%) trials. Risk of detection bias and attrition bias was judged to be low in the majority of the studies (59% and 71%, respectively). With regard to the risk of bias domains addressing a clustered design (relevant for n=45 included studies), low risk of bias was found in the majority of the studies (n=36 [80%] for recruitment bias, n=43 [96%] for unit of analysis error, and n=38 [84%] for baseline imbalances).

In 10 (20%) studies the overall risk of bias was judged to be low (see Additional file 3, table S1 and figure S2).

**Effectiveness of de-implementation strategies**
*De-implementation vs. usual care*
Of the 43 studies comparing de-implementation to usual care, 28 (65%) reported their de-implementation strategy to be successful: the targeted strategy significantly reduced the use of a healthcare practice compared to the usual care group. Success rates for diagnostic and therapeutic healthcare practices were 73% (n=8 of 11 studies) and 63% (n=20 of 32 studies), respectively.

Thirty of the 43 studies reported the relative change from baseline in the use of a healthcare practice for both the intervention and usual care groups or provided data to calculate this. In more than half of the studies (n=16) a reduction in both study groups was seen (lower left quadrant in Figure 2).

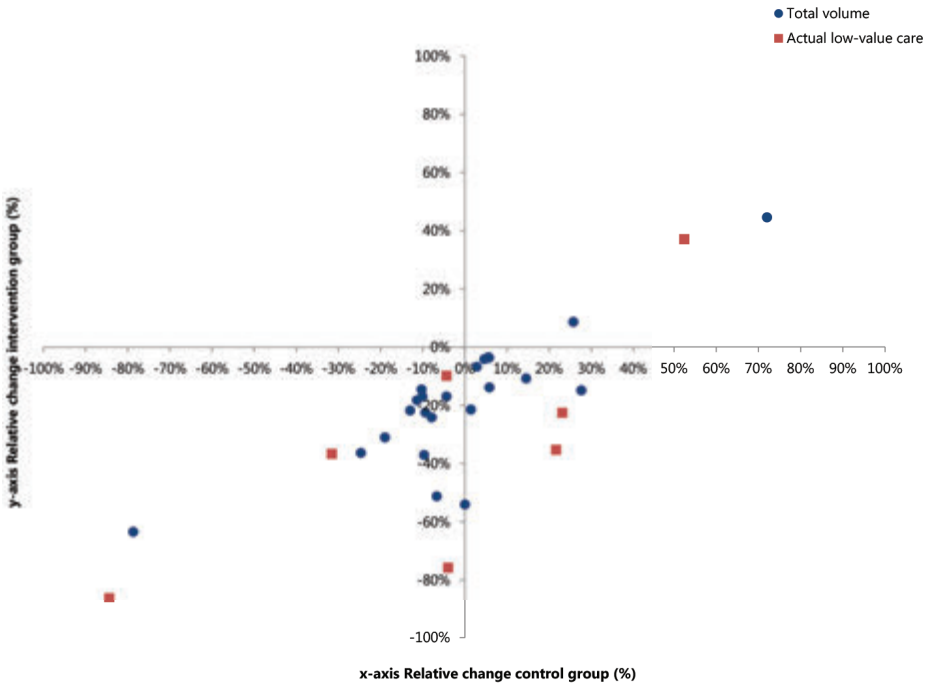



**Figure 2. Relative change in volume of health care use after de-implementation in both the intervention group and the control group**



The overall median difference in relative reductions between intervention and usual care groups was 13% (IQR 9% - 27%). Subsets of studies measuring actual low-value care (rather than total volume) (n=7) and studies at overall low risk of bias (n=6) showed a median difference of 15% (IQR 5% - 51%) and 16% (IQR 15% – 24%), respectively.

The median difference was 15% (IQR 9% - 21%) in studies addressing diagnostic healthcare practices and 12% (IQR 7% - 39%) in studies addressing therapeutic healthcare practices (Table 3). With regard to characteristics of de-implementation strategies in relation to the effect, the smallest median difference was found for strategies addressing a single target with a single intervention (n=5; 5%, IQR 5% – 20%). Multifaceted strategies showed a trend towards a larger effect compared to single intervention strategies, although three or more interventions in a strategy did not lead to a larger median difference compared to two interventions (13% [IQR 9% - 26%] vs. 14% [IQR 9% - 25%]). For strategies aiming to reduce therapeutic low-value care, patient-targeted strategies tend to achieve a larger median difference in relative reductions (20% [IQR 10% - 43%]) compared to strategies were no patients were addressed (median 11% [IQR 7% - 21%]). Strategies containing audit and feedback

had a larger median difference than strategies without this intervention (16% [IQR 9% - 27%] vs. 8% [IQR 6% - 13%]). Incorporating reminders seemed beneficial for strategies addressing diagnostic healthcare practices (median of 20% [10% - 25%] compared to 13% [11% - 15%] for strategies without reminders). Whether strategies were tailor-made based on pre-identified barriers and facilitators did not influence the effect.

**Table 3. Difference in relative reductions between de-implementation and usual care**

| | All studies | | Diagnostic healthcare practices | | Therapeutic healthcare practices | |
|---|---|---|---|---|---|---|
| | **n** | **Median (IQR)** | **n** | **Median (IQR)** | **n** | **Median (IQR)** |
| **De-implementation strategy** | 30 | 13% (9% - 27%) | 8 | 15% (9% - 21%) | 22 | 12% (7% - 39%) |
| ***Single or multiple targets and interventions*** | | | | | | |
| Single target, single intervention | 5 | 5% (5% - 20%) | 2 | 13% (9% - 16%) | 3 | 5% (3% - 25%) |
| Single target, combination of interventions | 16 | 16% (9% - 26%) | 5 | 16% (13% - 25%) | 11 | 15% (9% - 54%) |
| Multiple targets, combination of interventions | 9 | 12% (9% - 28%) | 1 | 10% (NA) | 8 | 12% (8% - %) |
| ***Number of intervention categories**** | | | | | | |
| 1 | 5 | 5% (5% - 20%) | 2 | 13% (9% - 16%) | 3 | 5% (3% - 25%) |
| 2 | 10 | 14% (9% - 25%) | 2 | 13% (11% - 14%) | 8 | 15% (9% - 35%) |
| 3 or more | 15 | 13% (9% - 26%) | 4 | 19% (12% - 26%) | 11 | 13% (7% - 42%) |
| ***Strategy targets*** | | | | | | |
| Provider only | 20 | 14% (8% - 23%) | 7 | 16% (11% - 22%) | 13 | 12% (7% - 23%) |
| Patient only | 1 | 46% (NA) | 0 | NA | 1 | 46% (NA) |
| Provider and patient | 5 | 28% (12% - 42%) | 0 | NA | 5 | 28% (12% - 42%) |
| Provider and organizational context | 1 | 9% (NA) | 0 | NA | 1 | 9% (NA) |
| Provider, patient, and organizational context | 2 | 10% (8% - 11%) | 0 | NA | 2 | 10% (8% - 11%) |
| Provider, patient, and healthcare system | 1 | 10% (NA) | 1 | 10% (NA) | 0 | NA |
| ***Interventions*** | | | | | | |
| Targeted at provider | 29 | 13% (9% - 25%) | 8 | 15% (9% - 21%) | 21 | 12% (7% - 28%) |
| Not targeted at provider | 1 | 46% (NA) | 0 | NA | 1 | 46% (NA) |
| Any education (either meetings or material, or both) | 24 | 13% (9% - 26%) | 6 | 15% (10% - 23%) | 18 | 12% (9% - 38%) |

| | | | | | | |
|---|---|---|---|---|---|---|
| No education | 6 | 13% (5% - 26%) | 2 | 13% (9% - 16%) | 4 | 16% (4% - 32%) |
| Reminders | 8 | 11% (8% - 21%) | 5 | 20% (10% - 25%) | 3 | 9% (5% - 10%) |
| No reminders | 22 | 14% (9% - 39%) | 3 | 13% (11% - 15%) | 19 | 15% (8% - 43%) |
| Audit and feedback | 22 | 16% (9% - 27%) | 5 | 16% (13% - 25%) | 17 | 15% (9% - 42%) |
| No audit and feedback | 8 | 8% (6% - 13%) | 3 | 10% (7% - 15%) | 5 | 7% (7% - 12%) |
| Targeted at patient | 9 | 13% (10% - 42%) | 1 | 10% (NA) | 8 | 20% (10% - 43%) |
| Not targeted at patient | 21 | 13% (9% - 23%) | 7 | 16% (11% - 22%) | 14 | 11% (7% - 21%) |
| Targeted at context | 3 | 9% (8% - 11%) | 0 | NA | 3 | 9% (8% - 11%) |
| Not targeted at context | 27 | 15% (9% - 27%) | 8 | 15% (9% - 21%) | 19 | 15% (8% - 43%) |
| Targeted at system | 1 | 10% (NA) | 1 | 10% (NA) | 0 | NA |
| Not targeted at system | 29 | 13% (9% - 27%) | 7 | 16% (11% - 22%) | 22 | 12% (7% - 39%) |
| ***Barriers & facilitators*** | | | | | | |
| Pre-identified | 3 | 12% (11% - 14%) | 1 | 10% (NA) | 2 | 14% (13% - 14%) |
| Not pre-identified | 27 | 13% (8% - 27%) | 7 | 16% (11% - 22%) | 20 | 12% (7% - 43%) |

NA=not available due to no or low number of studies
* Based on taxonomy provided by the Cochrane Effective Practice and Organisation of Care (EPOC) Group (see supplementary material)

Thirteen studies did not report the relative change from baseline in the use of a healthcare practice or data to calculate this, but rather presented (adjusted) odds ratios, hazard ratios, or mean differences (see Additional file 3, table S3). One study only addressed the sustainability of the de-implementation strategy and did not present results regarding the use of a healthcare practice before and shortly after de-implementation (i.e. post-intervention). For the other 12 studies, it is difficult to compare effect sizes because of the heterogeneity in outcome measures (e.g. odds ratios, hazard ratios). We present their results narratively instead: of the four studies addressing a single target with a single intervention, two reported a significant effect. Multifaceted strategies with two interventions (n=4 studies) did not lead to significant differences compared to usual care, whereas all four studies using three intervention categories reported a significant effect in favour of the de-implementation strategy with regard to reducing low-value care.

*Direct comparison of de-implementation strategies*

Eight studies compared a de-implementation strategy with another de-implementation strategy and reported the relative change from baseline in the use of a low-value healthcare practice or presented data to calculate this. These studies included various (combinations of) interventions, which resulted in 19 possible direct comparisons of strategies with little overlap between studies (see Additional file 3, table S4). Fourteen comparisons addressed interventions only targeted at healthcare providers. Overall, strategies with more interventions led to a larger relative reduction. The combination

of education and audit and feedback seemed successful. In one study the effect of an intervention targeted at providers (audit and feedback) was directly compared with a patient directed intervention and the latter led to a larger relative reduction. Two of the three strategies targeted at both provider and patient led to larger relative reductions compared to a strategy targeted at either one of these targets.

**Sustainability of effect**

Sustainability of the effect was assessed in five of the 49 studies (10%). Follow-up ranged from three to 18 months and two studies addressed diagnostic low-value care. One study did not report sustainability of results in enough detail, the other four studies reported a continued reduction in low-value care use. In one study, that aimed to reduce antibiotic prescribing with an intervention targeted at providers, the difference compared to usual care was statistically significant at three months follow-up. Two other studies that both aimed to reduce medication use and both targeted providers (and one patients as well), found no significant difference compared to usual care at three and 12 months. For the fourth study (addressing diagnostic low-value care), it was unclear whether the difference between the intervention targeted at providers and usual care was significant at 18 months follow-up.

## Discussion

In this systematic review, we included 49 RCTs that evaluated a strategy aimed at reducing low-value care. Over two third of the studies addressed the reduction of medication use and were performed in a primary care or outpatient setting. Compared to usual care, de-implementation strategies were successful in 65%, with an overall median relative reduction in the use of low-value healthcare practices of 13% (IQR 9% - 27%). The effect of de-implementation tended to be smaller for the subgroup of strategies consisting of a single intervention. To reduce therapeutic low-value care services, a strategy targeted at patients was inclined to achieve a larger effect compared to strategies that did not address patients. The subgroup of strategies containing audit and feedback showed a trend towards a larger effect than strategies without this intervention. Incorporating reminders seemed beneficial for strategies addressing diagnostic healthcare practices.

**Comparison with other studies**

Our findings confirm the results of Colla et al., who concluded that multicomponent interventions are potentially more effective in reducing low-value care than single-component interventions, especially when addressing both patients and clinicians.[2] Our results also show that multifaceted strategies have greater potential to reduce low-value

care practices. Yet, we furthermore found that effectiveness did not increase with the number of intervention categories (2 vs. 3 or more) in a multifaceted strategy. A similar result was reported in a systematic review of the effectiveness of guideline implementation strategies.[26] In general, in the literature on guideline implementation and behavioural change interventions there is no consensus regarding the number of interventions in relation to effectiveness. Some reviews did find multifaceted strategies to be more effective than strategies consisting of a single intervention,[27-30] whereas others did not.[26,31-34]

In the overview by Colla et al. the most effective interventions for de-implementation were clinical decision support tools, education, and patient education; and performance feedback was considered to be a promising strategy.[2] By quantifying the effect, we found a clear trend towards a larger effect for strategies incorporating an audit and feedback intervention, and reminders and decision support tools seemed beneficial in particular for de-implementation strategies addressing diagnostic low-value care. The potential of audit and feedback as an effective strategy to change behaviour is also known from the literature on implementation.[35,36] With regard to using reminders for implementation, mixed effectiveness was reported across various settings and circumstances, however, there was no evidence of specific reminder or contextual characteristics to be related to the degree of effect.[35 37,38]



It has been suggested that for changing behaviour it is not the number or type of interventions that matters, but the fact that an implementation strategy is context-specific and addresses existing barriers and facilitators to change.[3,39-41] De-implementation as well as implementation are intended to change behaviour, although changing existing care is likely to face different challenges than implementation of new practices.[12,42-44] It is therefore unfortunate that we identified only few studies that reported how assessment of barriers and facilitators informed the design of their de-implementation strategy.

**Strengths and limitations of this study**

The strength of our review is that we applied rigorous and systematic methods to explore the field of de-implementation research, including a rigorous assessment of the risk of bias in the included studies. In contrast to previous reviews, we only included randomised studies and quantified and compared the effectiveness of de-implementation strategies, preferably based on actual low-value care rather than total volume.

Despite our systematic search strategy, it is still possible that we missed relevant publications, because of the many different terms that are used to describe the process

of reducing low-value care.[15] Furthermore, our focus on RCTs might be an explanation for identifying almost no studies addressing financial or regulatory interventions targeted at the healthcare system. Randomised study designs are probably not the first choice when evaluating this type of - potentially effective - interventions. Nonetheless, we believe that, overall, our set of included studies is a representative sample of the existing evidence regarding strategies to reduce low-value care that can be developed and carried out in individual hospitals or healthcare organisations.

There were many different combinations of interventions used to reduce low-value care, with little overlap. As a result, we were faced with substantial heterogeneous de-implementation strategies and were not able to disentangle the effect of a single component.

Unfortunately, it was not possible to include all studies in our quantitative analysis, due to lack of available data and heterogeneity in outcome measures (i.e., absolute numbers, proportions, ratios, rates). Using relative changes between baseline and post-intervention to calculate effectiveness enabled us to include as much of the studies as possible (30 out of 49 RCTs). As a consequence of using median and interquartile ranges to summarize across studies, all studies had an equal weight in our analysis, which is different from conventional meta-analysis. The advantage of using the median rather than the mean, is that extreme results are less likely to influence the summary estimate.[37] Despite these challenges regarding heterogeneity and analyses, we are confident that our quantitative summary of 30 RCTs, complemented by the qualitative results, can support those who are planning to develop a de-implementation strategy.

**Implications for practice and areas for future research**

Most de-implementation initiatives seem to focus on reducing medication use, thereby targeting the healthcare provider. Based on our findings, it seems worthwhile to also target patients, as patient-provider interactions influence clinical decision making, and patient expectations and requests are among the barriers to change perceived by clinicians.[45-47] Informed and engaged patients will facilitate patient-provider communication, resulting in better motivation, satisfaction, and improved health outcomes.[48]

Knowledge on existing barriers and facilitators can inform the design of a de-implementation strategy, not only with regard to potential targets, but also with regard to the choice of interventions. If lack of knowledge is not the main driver for ongoing use of a particular low-value care, just providing education is unlikely to be effective. In that case an (additional) intervention addressing the healthcare provider's motivation, like audit and feedback, would have more potential. Currently, there is a

lack of evidence on how to optimally tailor strategies to specific contexts and what effect may be achieved.[41]

We found that a considerable reduction of low-value care was possible (median relative reduction of 13%), however, it will depend on the baseline level of low-value care and the context whether the actual impact of a strategy is clinically meaningful. A reduction of 13% might be insignificant for one low-value care practice, yet it could mean a substantial improvement of quality of care in other practices (e.g. when serious adverse events are prevented). In addition, the ultimate aim of de-implementation is a permanent reduction of a low-value care practice, however, only a minority of the included studies addressed sustainability.

It is essential that future studies on the topic provide all essential contextual information needed to interpret and apply their results, including knowledge on barriers and facilitators, sustainability of effect and insight into (unintended) consequences of reducing a low-value care practice on patient health or healthcare use. Authors of de-implementation studies should thereby use the relevant guidelines aimed at structured and transparent reporting, such as the Standards for Reporting Implementation Studies (StaRI) Statement, the Standards for Quality Improvement Reporting Excellence (SQUIRE 2.0), and the Template for Intervention Description and Replication (TIDieR) checklist and guide. [49-51]



## Conclusions

The majority of active de-implementation strategies identified by our systematic review were successful in reducing low value care, achieving a median relative reduction of 13%. These results should encourage healthcare professionals and policymakers to initiate their own de-implementation projects. Based on our findings, they are recommended to develop tailored, multifaceted de-implementation strategies and to consider audit and feedback and patient directed interventions. Our results strengthen the evidence-base to design successful de-implementation strategies. Insights into intervention details, sustainability of effects, and impact on health outcomes will further advance our understanding regarding the optimal approach to reduce low-value care.

## Acknowledgements

We thank René Spijker for developing and performing the searches, Toshihiko Takada for his assistance with study selection and data extraction, and Johannes B Reitsma for his contribution to planning the data analyses and drafting the results section of this manuscript.

## References


1. Scott IA, Duckett SJ. In search of professional consensus in defining and reducing low-value care. *Med J Aust.* 2015;203(4):179-181.
2. Colla CH, Mainor AJ, Hargreaves C, Sequist T, Morden N. Interventions Aimed at Reducing Use of Low-Value Health Services: A Systematic Review. *Med Care Res Rev.* 2017;74(5).
3. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet.* 2003;362(9391):1225-1230.
4. Morgan DJ, Brownlee S, Leppin AL, Kressin N, Dhruva SS, et al. Setting a research agenda for medical overuse. *BMJ.* 2015;351:h4534.
5. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, et al. Evidence for overuse of medical services around the world. *Lancet.* 2017.
6. Korenstein D, Falk R, Howell EA, Bishop T, Keyhani S. Overuse of health care services in the United States: an understudied problem. *Arch Intern Med.* 2012;172(2):171-178.
7. Levinson W, Kallewaard M, Bhatia RS, Wolfson D, Shortt S, Kerr EA. 'Choosing Wisely': a growing international campaign. *BMJ Qual Saf.* 2015;24(2):167-174.
8. Garner S, Littlejohns P. Disinvestment from low value clinical interventions: NICEly done? *BMJ.* 2011;343:d4519.
9. Elshaug AG, Watt AM, Mundy L, Willis CD. Over 150 potentially low-value health care practices: an Australian study. *Med J Aust.* 2012;197:556-560.
10. Wammes JJ, van den Akker-van Marle ME, Verkerk EW, van Dulmen SA, Westert GP, et al. Identifying and prioritizing lower value services from Dutch specialist guidelines and a comparison with the UK do-not-do list. *BMC Med.* 2016;14(1):196.
11. Rosenberg A, Agiro A, Gottlieb M, Barron J, Brady P, et al. Early Trends Among Seven Recommendations From the Choosing Wisely Campaign. *JAMA Intern Med.* 2015;175(12):1913-1920.
12. van Bodegom-Vos L, Davidoff F, Marang-van de Mheen PJ. Implementation and de-implementation: two sides of the same coin? *BMJ Qual Saf.* 2016.
13. Shojania KG, Grimshaw JM. Evidence-based quality improvement: the state of the science. *Health Aff (Millwood).* 2005;24(1):138-150.
14. NHS Centre for Reviews and Dissemination. Getting evidence into practice. *Effective Health Care.* 1999;5(1):1-16.
15. Niven DJ, Mrklas KJ, Holodinsky JK, Straus SE, Hemmelgarn BR, et al. Towards understanding the de-adoption of low-value clinical practices: A scoping review. *BMC Medicine.* 2015;13.
16. Grol R, Wensing M, Eccles MP, Davis D. *Improving patient care: the implementation of change in health care.*: John Wiley & Sons; 2013.
17. Effective Practice and Organisation of Care (EPOC); EPOC Taxonomy. *Available at: https://epoccochraneorg/epoc-taxonomy.* 2015.

18. *Effective Practice and Organisation of Care (EPOC); The EPOC taxonomy of health systems interventions. EPOC Resources for review authors*. Oslo: Norwegian Knowledge Centre for the Health Services;2016.
19. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
20. Higgins J, Green S, (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. In: The Cochrane Collaboration; 2011: www.handbook.cochrane.org.
21. Bolzern J, Mnyama N, Bosanquet K, Torgerson DJ. A review of cluster randomized trials found statistical evidence of selection bias. *J Clin Epidemiol*. 2018;99:106-112.
22. Brierley G, Brabyn S, Torgerson D, Watson J. Bias in recruitment to cluster randomized trials: a review of recent publications. *J Eval Clin Pract*. 2012;18(4):878-886.
23. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ*. 2003;327(7418):785-789.
24. *R: A Language and Environment for Statistical Computing* [computer program]. Vienna, Austria: R Foundation for Statistical Computing; 2019.
25. *Review Manager (RevMan)* [computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014.
26. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, et al. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess*. 2004;8:iii-iiv.
27. Grimshaw JM, Shirran L, Thomas R, Mowatt G, Fraser C, et al. Changing provider behavior: an overview of systematic reviews of interventions. *Med Care*. 2001;39:II2-45.
28. Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. *BMJ*. 1998;317(7156):465-468.
29. Wensing M, Grol R. Single and combined strategies for implementing changes in primary care: a literature review. *Int J Qual Health Care*. 1994;6(2):115-132.
30. van der Wees PJ, Jamtvedt G, Rebbeck T, de Bie RA, Dekker J, Hendriks EJ. Multifaceted strategies may increase implementation of physiotherapy clinical guidelines: a systematic review. *Aust J Physiother*. 2008;54(4):233-241.
31. Squires JE, Sullivan K, Eccles MP, Worswick J, Grimshaw JM. Are multifaceted interventions more effective than single-component interventions in changing health-care professionals' behaviours? An overview of systematic reviews. *Implement Sci*. 2014;9:152.

32. Sinuff T, Muscedere J, Adhikari NK, Stelfox HT, Dodek P, et al. Knowledge translation interventions for critically ill patients: a systematic review. *Crit Care Med.* 2013;41(11):2627-2640.
33. Grimshaw J, Eccles M, Tetroe J. Implementing clinical guidelines: current evidence and future implications. *J Contin Educ Health Prof.* 2004;24 Suppl 1:S31-37.
34. Wensing M, van der Weijden T, Grol R. Implementing guidelines and innovations in general practice: which interventions are effective? *Br J Gen Pract.* 1998;48(427):991-997.
35. Chan WV, Pearson TA, Bennett GC, Cushman WC, Gaziano TA, et al. ACC/AHA Special Report: Clinical Practice Guideline Implementation Strategies: A Summary of Systematic Reviews by the NHLBI Implementation Science Work Group: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J Am Coll Cardiol.* 2017;69(8):1076-1092.
36. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev.* 2012;6:CD000259.
37. Arditi C, Rege-Walther M, Durieux P, Burnand B. Computer-generated reminders delivered on paper to healthcare professionals: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev.* 2017;7:Cd001175.
38. Shojania KG, Jennings A, Mayhew A, Ramsay CR, Eccles MP, Grimshaw J. The effects of on-screen, point of care computer reminders on processes and outcomes of care. *Cochrane Database Syst Rev.* 2009(3):Cd001096.
39. Grimshaw JM, Eccles MP, Walker AE, Thomas RE. Changing physicians' behavior: what works and thoughts on getting more things to work. *J Contin Educ Health Prof.* 2002;22(4):237-243.
40. Gagliardi AR, Alhabib S. Trends in guideline implementation: a scoping systematic review. *Implement Sci.* 2015;10:54.
41. Baker R, Camosso-Stefinovic J, Gillies C, Shaw EJ, Cheater F, et al. Tailored interventions to address determinants of practice. *Cochrane Database Syst Rev.* 2015(4):Cd005470.
42. Roman BR, Asch DA. Faded promises: the challenge of deadopting low-value care. *Ann Intern Med.* 2014;161(2):149-150.
43. Scott IA, Elshaug AG. Foregoing low-value care: how much evidence is needed to change beliefs? *Intern Med J.* 2013;43(2):107-109.
44. Ubel PA, Asch DA. Creating value in health by understanding and overcoming resistance to de-innovation. *Health Aff (Millwood).* 2015;34(2):239-244.
45. Ackerman SL, Gonzales R, Stahl MS, Metlay JP. One size does not fit all: evaluating an intervention to reduce antibiotic prescribing for acute bronchitis. *BMC Health Serv Res.* 2013;13:462.

46. Mira JJ, Carrillo I, Silvestre C, Perez-Perez P, Nebot C, et al. Drivers and strategies for avoiding overuse. A cross-sectional study to explore the experience of Spanish primary care providers handling uncertainty and patients' requests. *BMJ Open*. 2018;8(6):e021339.
47. Osorio D, Ribera A, Solans-Domenech M, Arroyo-Moliner L, Ballesteros M, Romea-Lecumberri S. Healthcare professionals' opinions, barriers and facilitators towards low-value clinical practices in the hospital setting. *Gac Sanit*. 2019.
48. Riedl D, Schussler G. The Influence of Doctor-Patient Communication on Health Outcomes: A Systematic Review. *Z Psychosom Med Psychother*. 2017;63(2):131-150.
49. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, et al. Standards for Reporting Implementation Studies (StaRI) Statement. *BMJ*. 2017;356:i6795.
50. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf*. 2016;25(12):986-992.
51. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014;348:g1687.

## Supplementary material

### Search strategies

For data sources and search terms see Supplemental appendix of Chapter 6 (page 180).

### Classification of interventions

Based on taxonomy provided by the Cochrane Effective Practice and Organisation of Care (EPOC) Group[1,2]

| | |
|---|---|
| Educational meetings | Courses, workshops, conferences or other educational meetings<br>Including ***educational outreach visits*** (personal visits by a trained person to health workers in their own settings, to provide information with the aim of changing practice) and ***inter-professional education*** (continuing education for health professionals that involves more than one profession in joint, interactive learning) |
| Educational materials | Distribution to individuals, or groups, of educational materials to support clinical care, i.e., any intervention in which knowledge is distributed. For example this may be facilitated by the internet, learning critical appraisal skills; skills for electronic retrieval of information, diagnostic formulation; question formulation |
| Reminders | Manual or computerised interventions that prompt health workers to perform an action during a consultation with a patient, for example computer ***decision support systems***. |
| Audit and feedback | A summary of health workers' performance over a specified period of time, given to them in a written, electronic or verbal format. The summary may include recommendations for clinical action. |
| Patient directed interventions | Interventions aimed at patients; e.g. patient information, posters in waiting room, mass media campaign. |
| Organisational interventions | Interventions aimed at a group of professionals, interprovider relations, organisation or institution<br>Examples:<br>- Revision of professional roles: 'professional substitution', 'boundary encroachment'; includes the shifting of roles among health professionals. For example, nurse midwives providing obstetrical care; pharmacists providing drug counselling that was formerly provided by nurses and physicians; nutritionists providing nursing care; physical therapists providing nursing care. Also includes expansion of role to include new tasks.<br>- Clinical multidisciplinary teams: creation of a new team of health professionals of different disciplines or additions of new members to the team who work together to care for patients<br>- Formal integration of services across sectors or teams or the organisation of services to bring all services together at one time also sometimes called 'seamless care'<br>- Skill mix changes: changes in numbers, types or qualifications of staff |

| | |
|---|---|
| | - Local opinion leaders: the identification and use of identifiable local opinion leaders to promote good clinical practice.<br>- Continuity of care: including one or many episodes of care for inpatients or outpatients)<br>  • Arrangements for follow-up.<br>  • Case management (including co-ordination of assessment, treatment and arrangement for referrals<br>- Communication and case discussion between distant health professionals e.g. telephone links; telemedicine; there is a television/ video link between specialist and remote nurse practitioners<br>- Continuous quality improvement: an iterative process to review and improve care that includes involvement of healthcare teams, analysis of a process or system, a structured process improvement method or problem solving approach, and use of data analysis to assess changes<br>- Clinical Practice Guidelines: clinical guidelines are systematically developed statements to assist healthcare providers and patients to decide on appropriate health care for specific clinical circumstances'(US IOM).<br>- Clinical incident reporting: system for reporting critical incidents,<br>- Routine patient-reported outcome measures: routine administration and reporting of patient-reported outcome measures to providers and/or patients<br>- Local consensus processes: formal or informal local consensus processes, for example agreeing a clinical protocol to manage a patient group, adapting a guideline for a local health system or promoting the implementation of guidelines. |
| Structural interventions | - Changes to the setting/site of service delivery e.g. moving a family planning service from a hospital to a school<br>- Ownership, accreditation, and affiliation status of hospitals and other facilities |
| Regulatory interventions | Any intervention that aims to change health services delivery by regulation or law<br>Examples:<br>- Changes in medical liability<br>- Licensure |
| Financial interventions | Any financial interventions aimed at either healthcare professional, patient, health care system |

## Results

*Included studies*

**Table S1. Included studies (n=49)**

| Reference | Country | Unit of randomization | Low overall risk of bias* | Type of low-value care | Low-value care |
|---|---|---|---|---|---|
| Allard 2001[3] | Canada | Individual participant | No | Medication | General |
| Awad 2006[4] | Sudan | Healthcare centre / practice / group of providers | No | Medication | Antibiotics |
| Baker 2003[5] | UK | Healthcare centre / practice / group of providers | No | Diagnostic | Pathology tests |
| Bates 1999[6] | USA | Individual participant | No | Diagnostic | Laboratory tests |
| Bhatia 2014[7] | USA | Provider | Yes | Diagnostic | Imaging |
| Bhatia 2017[8] | Canada, USA | Provider | No | Diagnostic | Imaging |
| Briel 2006[9] | Switserland | Provider | No | Medication | Antibiotics |
| Carney 2012[10] | USA | Provider | No | Diagnostic | Screening |
| Clyne 2018[11,12] | Ireland | Healthcare centre / practice / group of providers | Yes | Medication | General |
| Coenen 2004[13] | Belgium | Provider | No | Medication | Antibiotics |
| Daley 2018[14] | Canada | Individual participant | No | Medication | Antibiotics |
| Davies 2002[15] | Canada | Healthcare centre / practice / group of providers | No | Non-medication | Electronic fetal monitoring |
| De Burgh 1995[16] | Australia | Provider | No | Medication | Benzodiazepines |
| Eccles 2001[17] | UK | Healthcare centre / practice / group of providers | No | Diagnostic | Imaging, referral |
| Fenton 2016[18] | USA | Provider | No | Diagnostic | Imaging |
| Fine 2003[19] | USA | Provider | No | Medication | Antibiotics |
| Finkelstein 2001[20] | USA | Healthcare centre / practice / group of providers | No | Medication | Antibiotics |

| Setting | Strategy target | Strategy – interventions** |
|---|---|---|
| Primary care, outpatient services | Provider | Educational material, Audit and feedback |
| Primary care, outpatient services | Provider | Educational meetings, Audit and feedback |
| Primary care, outpatient services | Provider | Educational material, Audit and feedback |
| Hospital | Provider | Reminders |
| Primary care, outpatient services | Provider | Educational meetings, Educational material, Audit and feedback |
| Primary care, outpatient services | Provider | Educational meetings, Educational material, Audit and feedback |
| Primary care, outpatient services | Provider | Educational meetings, Educational material |
| Other | Provider | Educational material, Audit and feedback |
| Primary care, outpatient services | Provider and patient | Educational meetings, Educational material, Patient directed interventions (educational material) |
| Primary care, outpatient services | Provider and patient | Educational meetings, Educational material, Reminders, Patient directed interventions (educational material as part of public campaign also including television spots, and radio messages) |
| Hospital | Organisation | Organisational interventions (modified report) |
| Hospital | Provider | Educational meetings, Educational material, Audit and feedback |
| Primary care, outpatient services | Provider | Educational meetings, Educational material |
| Primary care, outpatient services | Provider | Educational material, Reminders, Audit and feedback |
| Primary care, outpatient services | Provider | Educational meetings |
| Hospital | Provider | Educational material, Reminders |
| Primary care, outpatient services | Provider, patient and organisation | Educational meetings, Educational material, Patient directed interventions (educational materials via mail and in waiting rooms), Organisational interventions (peer leader) |

| | | | | | |
|---|---|---|---|---|---|
| Finkelstein 2008[21] | USA | Community | Yes | Medication | Antibiotics |
| Flottorp 2002[22] | Norway | Healthcare centre / practice / group of providers | No | Mixed | Antibiotics, laboratory tests, and clinical examination |
| Gjelstad 2013[23] | Norway | Healthcare centre / practice / group of providers | No | Medication | Antibiotics |
| Lemiengre 2018[24] | Belgium | Healthcare centre / practice / group of providers | No | Medication | Antibiotics |
| Linder 2009[25] | USA | Healthcare centre / practice / group of providers | No | Medication | Antibiotics |
| Loeb 2005[26] | Canada | Provider | Yes | Medication | Antibiotics |
| Mainous III 2000[27] | USA | Provider | No | Medication | Antibiotics |
| Meeker 2014[28] | USA | Provider | No | Medication | Antibiotics |
| Meeker 2016[29] | USA | Healthcare centre / practice / group of providers | Yes | Medication | Antibiotics |
| Metlay 2007[30] | USA | Community | No | Medication | Antibiotics |
| Monette 2007[31] | Canada | Healthcare centre / practice / group of providers | No | Medication | Antibiotics |
| Payne 1991[32] | USA | Healthcare centre / practice / group of providers | No | Non-medication | Resources |
| Persell 2016[33] | USA | Provider | No | | Antibiotics |
| Pettersson 2011[34] | Sweden | Healthcare centre / practice / group of providers | No | Medication | Antibiotics |
| Pimlott 2003[35] | Canada | Provider | No | Medication | Benzodiazepines |
| Rognstad 2013[36] | Norway | Healthcare centre / practice / group of providers | No | Medication | General |
| Samore 2005[37] | USA | Community | No | Medication | Antibiotics |

| | | |
|---|---|---|
| Primary care, outpatient services | Provider and patient | Educational meetings, Educational material, Audit and feedback, Patient directed interventions (educational material via letter, newsletters, posters, handouts website; training) |
| Primary care, outpatient services | Provider, patient and system | Educational meetings, Educational material, Reminders, Patient directed interventions (educational material in electronic and paper format), Financial interventions (increase in fee for telephone consultations) |
| Primary care, outpatient services | Provider and organisation | Educational meetings, Reminders, Audit and feedback, Organisational interventions (peer academic detailer; software tool for registration) |
| Primary care, outpatient services | Patient and organisation | Patient directed (eliciting parental concern and providing a safety net, information leaflet) and organisational (reducing clinicians' uncertainty with an objective inflammatory parameter) interventions |
| Primary care, outpatient services | Provider | Educational material, Reminders |
| Long term care facility | Provider | Educational meetings, Educational material, Reminders |
| Primary care, outpatient services | Provider and patient | Audit and feedback, Patient directed interventions (educational material) |
| Primary care, outpatient services | Patient | Patient directed interventions (poster-sized letters in examination rooms) |
| Primary care, outpatient services | Provider | Educational material, Reminders, Audit and feedback |
| Hospital | Provider, patient and organisation | Educational meetings, Educational material, Audit and feedback, Patient directed interventions (educational material in waiting and examination rooms; computerized education in waiting room) , Organisational interventions (clinical leaders) |
| Long term care facility | Provider | Educational material, Audit and feedback |
| Primary care, outpatient services | Provider | Audit and feedback |
| Primary care, outpatient services | Provider | Educational material, Reminders, Audit and feedback |
| Long term care facility | Provider | Educational meetings, Educational material, Audit and feedback |
| Primary care, outpatient services | Provider | Educational material, Audit and feedback |
| Primary care, outpatient services | Provider | Educational meetings, Audit and feedback |
| Primary care, outpatient services | Provider | Educational meetings, Educational material, Reminders |

| | | | | | |
|---|---|---|---|---|---|
| Seager 2005[38] | UK | Healthcare centre / practice / group of providers | No | Medication | Antibiotics |
| Shojania 1998[39] | USA | Provider | No | Medication | Antibiotics, vancomycin |
| Simon 2006[40] | USA | Healthcare centre / practice / group of providers | No | Medication | General |
| Tamblyn 2003[41] | Canada | Provider | No | Medication | General |
| Taylor 2005[42] | USA | Individual participant | No | Medication | Antibiotics |
| Thomas 2006[43] | Scotland (United Kingdom) | Healthcare centre / practice / group of providers | No | Diagnostic | Laboratory tests |
| Tierney 1990[44] | USA | Provider | No | Diagnostic | Test ordering |
| Trietsch 2017[45] | Netherlands | Healthcare centre / practice / group of providers | Yes | Diagnostic | Laboratory tests; prescribing |
| Urbiztondo 2017[46] | Argentina, Bolivia, Uruguay, and Paraguay | Healthcare centre / practice / group of providers | No | Medication | Antibiotics |
| van Driel 2007[47] | Belgium | Healthcare centre / practice / group of providers | No | Medication | Antibiotics |
| van Eijk 2001[48] | Netherlands | Healthcare centre / practice / group of providers | No | Medication | Antidepressants |
| Verstappen 2003[49] | Netherlands | Healthcare centre / practice / group of providers | Yes | Diagnostic | Test ordering |
| Voorn 2017[50] | Netherlands | Healthcare centre / practice / group of providers | Yes | Non-medication | Blood transfusion |
| Wei 2017[51] | China | Healthcare centre / practice / group of providers | Yes | Medication | Antibiotics |
| Welschen 2004[52] | Netherlands | Healthcare centre / practice / group of providers | Yes | Medication | Antibiotics |

RCT: randomized controlled trial
*Studies were classified as low risk of bias when they had 1) an adequate random sequence generation, 2) scored a low risk of bias for all three domains related to cluster randomised designs and 3) no high risk of bias due to unconcealed allocation, detection bias, attrition bias, or reporting bias, with unclear risk of bias for a maximum of two of these domains. ** Based on taxonomy provided by the Cochrane Effective Practice and Organisation of Care (EPOC) Group(1, 2)

| | | |
|---|---|---|
| Primary care, outpatient services | Provider | Educational meetings, Educational material |
| Hospital | Provider | Reminders |
| Health Maintenance Organisation | Provider | Educational meetings, Reminders |
| Primary care, outpatient services | Provider | Reminders |
| Pediatric practices | Patient | Patient directed interventions (educational material [video with accompanying pamphlet]) |
| Primary care, outpatient services | Provider | Educational material, Reminders, Audit and feedback |
| Primary care, outpatient services | Provider | Reminders, Audit and feedback |
| Primary care, outpatient services | Provider | Educational meetings, Audit and feedback |
| Primary care, outpatient services | Provide and patient | Educational meetings, Educational material, Audit and feedback, Patient directed interventions (educational material) |
| Primary care, outpatient services | Provider | Educational meetings, Educational material |
| Primary care, outpatient services | Provider | Educational meetings, Educational material, Audit and feedback |
| Primary care, outpatient services | Provider | Educational meetings, Educational material, Audit and feedback |
| Hospital | Provider | Educational meetings, Educational material, Audit and feedback |
| Primary care, outpatient services | Provide and patient | Educational meetings, Educational material, Audit and feedback, Patient directed interventions (information leaflet and video) |
| Primary care, outpatient services | Provider and patient | Educational meetings, Audit and feedback, Patient directed interventions (educational material) |

*De-implementation strategies*

**Table S2. De-implementation strategies in included studies**

| **Single target, single intervention (n=9)** | **n** | **Single target, combination of interventions (n=27)** | **n** | **Multiple targets, combination of interventions (n=13)** | **n** |
|---|---|---|---|---|---|
| **Provider** | **5** | **Provider** | **28** | **Provider and patient** | **8** |
| Reminders | 4 | Education (meetings & material) + audit and feedback | 7 | Education (meetings & material) + audit and feedback + patient directed intervention | 3 |
| Educational meetings | 1 | Educational material + audit and feedback | 5 | Education (meetings & material) + patient directed intervention | 2 |
| Audit and feedback | 1 | Educational material + reminders + audit and feedback | 4 | Education (meetings & material) + reminders + patient directed intervention | 1 |
| **Patient** | **2** | Education (meetings & material) | 3 | Educational meetings + audit and feedback + patient directed intervention | 1 |
| **Organisational context** | **1** | Educational meetings + audit and feedback | 3 | Audit and feedback + patient directed intervention | 1 |
| | | Education (meetings & material) + reminders | 2 | **Provider, patient, and organisational context** | **2** |
| | | Educational material + reminders | 2 | Education (meetings & material) + audit and feedback + patient directed + organisational | 1 |
| | | Educational meetings + reminders | 1 | Education (meetings & material) + patient directed + organisational | 1 |
| | | | | **Provider and organisational context** | **1** |
| | | | | Educational meetings + reminders + audit and feedback + organisation | 1 |
| | | | | **Patient and organisational context** | **1** |
| | | | | **Provider, patient, and system** | **1** |
| | | | | Education (meetings & material) + reminders + patient directed intervention + financial intervention | 1 |

*Effectiveness*

**Table S3. Results for studies not providing (data to calculate) the relative change from baseline in the use of a healthcare practice (n=13)**

| Study | Type of low-value care | Targeted at | Number of Intervention categories* | Intervention categories* | Outcome description | Results for intervention vs. control group, as reported by study authors |
|---|---|---|---|---|---|---|
| Fenton 2016 | Diagnostic | Provider | 1 | Educational meetings | Test ordering in intervention relative to control group | Adjusted OR1.07 (95%CI 0.49 to 2.32) |
| Shojania 1998 | Therapeutic | Provider | 1 | Reminders | Initiated or renewed order for vancomycin | 7.4 versus 10.3 orders per physician; p=0.02 |
| van Driel 2007 | Therapeutic | Provider | 2 | Education (meetings & material) | Prescribing first choice antibiotic | Adjusted OR 1.07 (95% CI 0.34 to 3.37) |
| Fine 2003 | Therapeutic | Provider | 2 | Educational material + reminders | Conversion to oral antibiotics | HR 1.23, (95%CI 1.00 to 1.52) |
| Linder 2009 | Therapeutic | Provider | 2 | Educational material + reminders | Antibiotic prescribing rate | Adjusted OR 0.8 (95%CI 0.6 to 1.2) |
| Bhatia 2014 | Diagnostic | Provider | 3 | Education (meetings & material) + audit and feedback | Ordering an appropriate test | OR 2.7 (95% CI 1.5 to 5.1) |
| Bhatia 2017 | Diagnostic | Provider | 3 | Education (meetings & material) + audit and feedback | Ordering a rarely appropriate test | OR 0.75 (95% CI 0.57 to 0.99) |
| Loeb 2005 | Therapeutic | Provider | 3 | Education (meetings & material) + reminders | Courses of antimicrobials for suspected urinary tract infections per 1000 resident days | Weighted mean difference – 0.49 (95%CI – 0.93 to – 0.06) |
| Taylor 2005 | Therapeutic | Patient | 1 | Patient directed intervention | Mean number of visits per patient where antibiotics were prescribed for otitis media | 1.7 (SD 2.1) vs. 1.9 (SD 2.4); p=0.23 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Daley 2018 | Therapeutic | Organisational context | 1 | Interventions directed at organisational context | Rate of treatment of asymptomatic bacteriuria | 10/35 (37.1%) vs. 24/41 (58.5%); p=0.016 |
| Seager 2005 | Therapeutic | Provider and patient | 3 | Education (meetings & material) + patient directed intervention | Inappropriate antibiotic prescriptions | OR 0.33 (95% 0.21 to 0.54) |
| Lemiengre 2018 | Therapeutic | Patient and organisational context | 2 | Interventions directed at patient and at organisational context | Antibiotic prescribing | Crude OR 1.07 (95% CI 0.61 to 1.88); adjusted OR 1.21 (95% CI 0.66 to 2.22) |
| *Only addressing sustainability, no results directly post intervention* | | | | | | |
| Clyne 2016 | Therapeutic | Provider and patient | 3 | Education (meetings & material) + patient directed intervention | New instances of potentially inappropriate prescribing | At 1 year post intervention 16 (in 12 [13 %] participants vs. 18 (in 18 [20 % participants]; p=0.38 |

* Based on taxonomy provided by the Cochrane Effective Practice and Organisation of Care (EPOC) Group[1,2]

**Table S4. Difference between relative reductions in use of a healthcare practice for direct comparison of de-implementation strategies (19 comparisons; n=8 studies)**

| Study | Type of low-value care | Number of intervention categories compared | Intervention categories compared | Difference in relative reductions from baseline |
|---|---|---|---|---|
| **Targeted at provider only** | | | | |
| Eccles 2001 | Diagnostic | 1 vs. 1 | Audit and feedback vs. reminders | -12.1% |
| Thomas 2006 | Diagnostic | 2 vs. 1 | Educational material + audit and feedback vs. reminders | 11.6% |
| Awad 2006 | Therapeutic | 2 vs. 1 | **Educational meetings (academic detailing)** + audit and feedback vs. Audit and feedback | 43.4% |
| Awad 2006 | Therapeutic | 2 vs. 1 | **Educational meetings (seminar)** + audit and feedback vs. Audit and feedback | 32.1% |
| Simon 2006 | Therapeutic | 2 vs. 1 | **Educational meetings** + reminders vs. reminders | 4.8% |
| Eccles 2001 | Diagnostic | 2 vs. 1 | Reminders + **audit and feedback** vs. reminders | 7.3% |
| Meeker 2016 | Therapeutic | 2 vs. 1 | Educational material + **audit and feedback** vs. educational material | 35.8% |
| Eccles 2001 | Diagnostic | 2 vs. 1 | **Reminders** + audit and feedback vs. audit and feedback | 19.4% |
| Meeker 2016 | Therapeutic | 2 vs. 1 | Educational material + **reminders (accountable justification)** vs. educational material | 31.9% |
| Meeker 2016 | Therapeutic | 2 vs. 1 | Educational material + **reminders (suggested alternatives)** vs. educational material | 26.8% |
| Thomas 2006 | Diagnostic | 3 vs. 2 | Educational material + **reminders** + audit and feedback vs. educational material + audit and feedback | -1.8% |
| Thomas 2006 | Diagnostic | 3 vs. 1 | **Educational material** + reminders + **audit and feedback** vs. reminders | 9.8% |

| | | | | |
|---|---|---|---|---|
| Awad 2006 | Therapeutic | 2 vs. 2 | Educational meetings (**academic detailing**) + audit and feedback vs. Educational meetings (**seminar**) + audit and feedback | 11.3% |
| Van Eijk 2001 | Therapeutic | 3 vs. 3 | Educational meetings (**individual visit**) + educational material + audit and feedback vs. Educational meetings (**group visit**) + educational material + audit and feedback | -29.8% |
| **Targeted at either providers or patients** | | | | |
| Mainous III 2000 | Therapeutic | 1 vs. 1 | Audit and feedback vs. patient directed interventions | -14.0% |
| **Targeted both providers and patients** | | | | |
| Mainous III 2000 | Therapeutic | 2 vs. 1 | **Audit and feedback** + patient directed interventions vs. patient directed interventions | -5.0% |
| Urbiztondo 2017 | Therapeutic | 4 vs. 3 | Educational meetings + Educational material + **audit and feedback** + patient directed interventions vs. Educational meetings + Educational material + patient directed interventions | 6.92% |
| Mainous III 2000 | Therapeutic | 2 vs. 1 | Audit and feedback + **patient directed interventions** vs. audit and feedback | 9.0% |
| Samore 2005 | Therapeutic | 4 vs. 1 | **Educational meetings + educational material + reminders** + patient directed interventions vs. patient directed interventions | 10.5% |

*Risk of bias assessment*

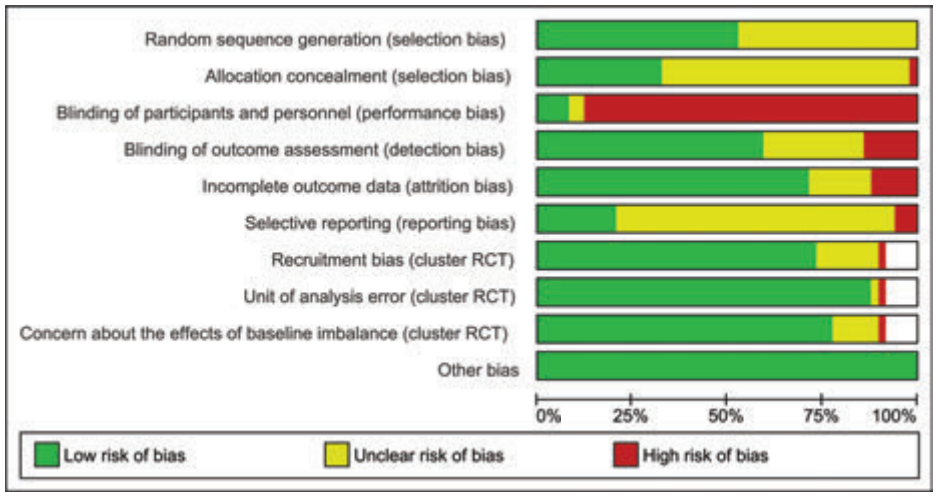
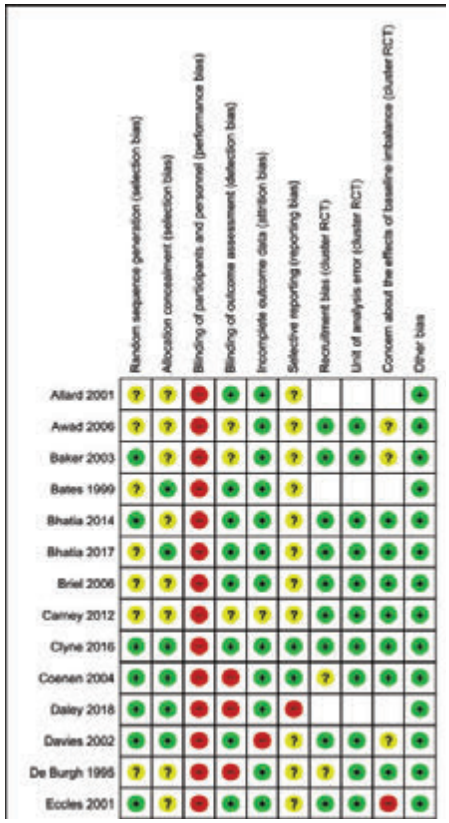



**Figure S1. Risk of bias graph**



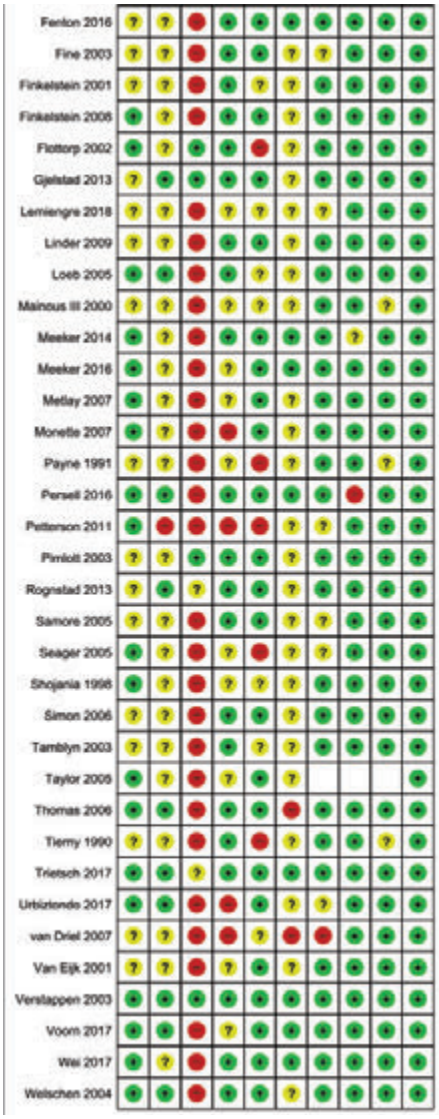



**Figure S2. Risk of bias summary**

**References**

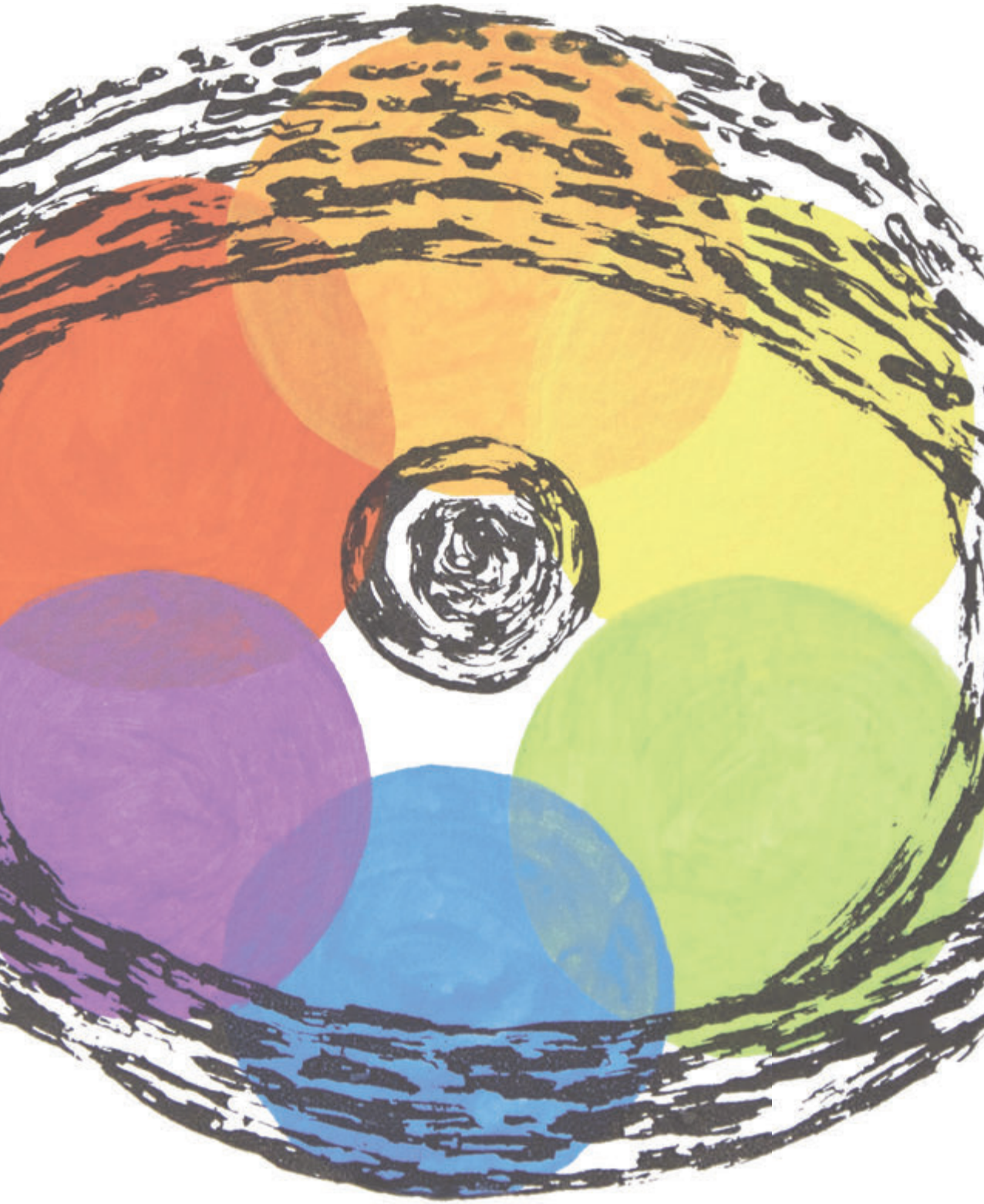

1. Effective Practice and Organisation of Care (EPOC); EPOC Taxonomy. *Available at: https://epoccochraneorg/epoc-taxonomy.* 2015.
2. *Effective Practice and Organisation of Care (EPOC); The EPOC taxonomy of health systems interventions. EPOC Resources for review authors.* Oslo: Norwegian Knowledge Centre for the Health Services;2016.
3. Allard J, Hébert R, Rioux M, Asselin J, Voyer L. Efficacy of a clinical medication review on the number of potentially inappropriate prescriptions prescribed for community-dwelling elderly people. *CMAJ.* 2001;164(9):1291-1296.
4. Awad AI, Eltayeb IB, Baraka OZ. Changing antibiotics prescribing practices in health centers of Khartoum State, Sudan. *Eur J Clin Pharmacol.* 2006;62(2):135-142.
5. Baker R, Falconer Smith J, Lambert Paul C. Randomised controlled trial of the effectiveness of feedback in improving test ordering in general practice. *Scand J Prim Health Care.* 2003;21(4):219-223.
6. Bates DW, Kuperman GJ, Rittenberg E, Teich JM, Fiskio J, et al. A randomized trial of a computer-based intervention to reduce utilization of redundant laboratory tests. *Am J Med.* 1999;106(2):144-150.
7. Bhatia RS, Dudzinski DM, Malhotra R, Milford CE, Sanborn DMY, et al. Educational intervention to reduce outpatient inappropriate echocardiograms: A randomized control trial. *JACC Cardiovasc Imaging.* 2014;7:857-866.
8. Bhatia RS, Ivers NM, Yin XC, Myers D, Nesbitt GC, et al. Improving the Appropriate Use of Transthoracic Echocardiography: The Echo WISELY Trial. *J Am Coll Cardiol.* 2017;70(9):1135-1144.
9. Briel M, Langewitz W, Tschudi P, Young J, Hugenschmidt C, Bucher Heiner C. Communication training and antibiotic use in acute respiratory tract infections. A cluster randomised controlled trial in general practice. *Swiss medical weekly.* 2006;136(15-16):241-247.
10. Carney PA, Abraham L, Cook A, Feig SA, Sickles EA, et al. Impact of an educational intervention designed to reduce unnecessary recall during screening mammography. *Acad Radiol.* 2012;19:1114-1120.
11. Clyne B, Smith SM, Hughes CM, Boland F, Bradley MC, et al. Effectiveness of a Multifaceted Intervention for Potentially Inappropriate Prescribing in Older Patients in Primary Care: A Cluster-Randomized Controlled Trial (OPTI-SCRIPT Study). *Ann Fam Med.* 2015;13(6):545-553.
12. Clyne B, Smith SM, Hughes CM, Boland F, Cooper JA, Fahey T. Sustained effectiveness of a multifaceted intervention to reduce potentially inappropriate prescribing in older patients in primary care (OPTI-SCRIPT study). *Implement Sci.* 2016;11(1):79.

13. Coenen S, Van Royen P, Michiels B, Denekens J. Optimizing antibiotic prescribing for acute cough in general practice: a cluster-randomized controlled trial. *J Antimicrob Chemother.* 2004;54(3):661-672.
14. Daley P, Garcia D, Inayatullah R, Penney C, Boyd S. Modified Reporting of Positive Urine Cultures to Reduce Inappropriate Treatment of Asymptomatic Bacteriuria Among Nonpregnant, Noncatheterized Inpatients: A Randomized Controlled Trial. *Infect Control Hosp Epidemiol.* 2018;39(7):814-819.
15. Davies B, Hodnett E, Hannah M, O'Brien-Pallas L, Pringle D, et al. Fetal health surveillance: a community-wide approach versus a tailored intervention for the implementation of clinical practice guidelines. *CMAJ.* 2002;167(5):469-474.
16. de Burgh, Mant A, Mattick RP, Donnelly N, Hall W, Bridges-Webb C. A controlled trial of educational visiting to improve benzodiazepine prescribing in general practice. *Aust J Public Health.* 1995;19:142-148.
17. Eccles M, Steen N, Grimshaw J, Thomas L, McNamee P, et al. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. *Lancet.* 2001;357(9266):1406-1409.
18. Fenton JJ, Kravitz RL, Jerant A, Paterniti DA, Bang H, et al. Promoting Patient-Centered Counseling to Reduce Use of Low-Value Diagnostic Tests: A Randomized Clinical Trial. *JAMA Intern Med.* 2016;176(2):191-197.
19. Fine MJ, Stone RA, Lave JR, Hough LJ, Obrosky DS, et al. Implementation of an evidence-based guideline to reduce duration of intravenous antibiotic therapy and length of stay for patients hospitalized with community-acquired pneumonia: a randomized controlled trial. *Am J Med.* 2003;115(5):343-351.
20. Finkelstein JA, Davis RL, Dowell SF, Metlay JP, Soumerai SB, et al. Reducing antibiotic use in children: a randomized trial in 12 practices. *Pediatrics.* 2001;108:1-7.
21. Finkelstein JA, Huang SS, Kleinman K, Rifas-Shiman SL, Stille CJ, et al. Impact of a 16-community trial to promote judicious antibiotic use in Massachusetts. *Pediatrics.* 2008;121(1):e15-23.
22. Flottorp S, Oxman AD, Havelsrud K, Treweek S, Herrin J. Cluster randomised controlled trial of tailored interventions to improve the management of urinary tract infections in women and sore throat. *BMJ.* 2002;325(7360):367.
23. Gjelstad S, Høye S, Straand J, Brekke M, Dalen I, Lindbæk M. Improving antibiotic prescribing in acute respiratory tract infections: cluster randomised trial from Norwegian general practice (prescription peer academic detailing (Rx-PAD) study). *BMJ.* 2013;347:f4403-f4403.
24. Lemiengre MB, Verbakel JY, Colman R, De Burghgraeve T, Buntinx F, et al. Reducing inappropriate antibiotic prescribing for children in primary care: a cluster randomised controlled trial of two interventions. *Br J Gen Pract.* 2018;68(668):e204-e210.

25. Linder JA, Schnipper JL, Tsurikova R, Yu T, Volk LA, et al. Documentation-based clinical decision support to improve antibiotic prescribing for acute respiratory infections in primary care: a cluster randomised controlled trial. *Inform Prim Care*. 2009;17(4):231-240.
26. Loeb M, Brazil K, Lohfeld L, McGeer A, Simor A, et al. Effect of a multifaceted intervention on number of antimicrobial prescriptions for suspected urinary tract infections in residents of nursing homes: cluster randomised controlled trial. *BMJ*. 2005;331:669.
27. Mainous Iii AG, Hueston WJ, Love MM, Evans ME, Finger R. An evaluation of statewide strategies to reduce antibiotic overuse. *Fam Med*. 2000;32:22-29.
28. Meeker D, Knight TK, Friedberg MW, Linder JA, Goldstein NJ, et al. Nudging guideline-concordant antibiotic prescribing: A randomized clinical trial. *JAMA Intern Med*. 2014;174:425-431.
29. Meeker D, Linder JA, Fox CR, Friedberg MW, Persell SD, et al. Effect of Behavioral Interventions on Inappropriate Antibiotic Prescribing Among Primary Care Practices: A Randomized Clinical Trial. *JAMA*. 2016;315(6):562-570.
30. Metlay JP, Camargo, Jr., C A, MacKenzie T, McCulloch C, et al. Cluster-Randomized Trial to Improve Antibiotic Use for Adults With Acute Respiratory Infections Treated in Emergency Departments. *Ann Emerg Med*. 2007;50:221-230.
31. Monette J, Miller MA, Monette M, Laurier C, Boivin JF, et al. Effect of an educational intervention on optimizing antibiotic prescribing in long-term care facilities. *J Am Geriatr Soc*. 2007;55:1231-1235.
32. Payne SMC, Ash A, Restuccia JD. The role of feedback in reducing medically unnecessary hospital use. *Med Care*. 1991;29:91-106.
33. Persell SD, Doctor JN, Friedberg MW, Meeker D, Friesema E, et al. Behavioral interventions to reduce inappropriate antibiotic prescribing: a randomized pilot trial. *BMC Infect Dis*. 2016;16:373.
34. Pettersson E, Vernby A, Molstad S, Lundborg CS. Can a multifaceted educational intervention targeting both nurses and physicians change the prescribing of antibiotics to nursing home residents? A cluster randomized controlled trial. *J Antimicrob Chemother.* 2011;66:2659-2666.
35. Pimlott NJG, Hux JE, Wilson LM, Kahan M, Li C, Rosser WW. Educating physicians to reduce benzodiazepine use by elderly patients: a randomized controlled trial. *CMAJ*. 2003;168(7):835-839.
36. Rognstad S, Brekke M, Fetveit A, Dalen I, Straand J. Prescription peer academic detailing to reduce inappropriate prescribing for older patients: A cluster randomised controlled trial. *Br J Gen Pract*. 2013;63:e554-e562.
37. Samore MH, Bateman K, Alder SC, Hannah E, Donnelly S, et al. Clinical decision support and appropriateness of antimicrobial prescribing: A randomized trial. *JAMA*. 2005;294:2305-2314.

38. Seager JM, Howell-Jones RS, Dunstan FD, Lewis MAO, Richmond S, Thomas DW. A randomised controlled trial of clinical outreach education to rationalise antibiotic prescribing for acute dental pain in the primary care setting. *Br Dent J*. 2005;201(4):217-222; discussion 216.
39. Shojania KG, Yokoe D, Platt R, Fiskio J, Ma'luf N, Bates DW. Reducing vancomycin use utilizing a computer guideline: results of a randomized controlled trial. *J Am Med Inform Assoc*. 1998;5:554-562.
40. Simon SR, Smith DH, Feldstein AC, Perrin N, Yang X, et al. Computerized prescribing alerts and group academic detailing to reduce the use of potentially inappropriate medications in older people. *J Am Geriatr Soc*. 2006;54:963-968.
41. Tamblyn R, Huang A, Perreault R, Jacques A, Roy D, et al. The medical office of the 21st century (MOXXI): effectiveness of computerized decision-making support in reducing inappropriate prescribing in primary care. *CMAJ*. 2003;169(6):549-556.
42. Taylor J, A., Kwan-Gett TSC, McMahon EM. Effectiveness of a parental educational intervention in reducing antibiotic use in children: a randomized controlled trial. *Pediatr Infect Dis J*. 2005;24(6):489-493.
43. Thomas RE, Croal BL, Ramsay C, Eccles M, Grimshaw J. Effect of enhanced feedback and brief educational reminder messages on laboratory test requesting in primary care: a cluster randomised trial. *Lancet*. 2006;367(9527):1990-1996.
44. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. *N Engl J Med*. 1990;322(21):1499-1504.
45. Trietsch J, van Steenkiste B, Grol R, Winkens B, Ulenkate H, et al. Effect of audit and feedback with peer review on general practitioners' prescribing and test ordering performance: a cluster-randomized controlled trial. *BMC Fam Pract*. 2017;18(1):53.
46. Urbiztondo I, Bjerrum L, Caballero L, Suarez MA, Olinisky M, Cordoba G. Decreasing Inappropriate Use of Antibiotics in Primary Care in Four Countries in South America-Cluster Randomized Controlled Trial. *Antibiotics (Basel)*. 2017;6(4).
47. van Driel M, L., Coenen S, Dirven K, Lobbestael J, Janssens I, et al. What is the role of quality circles in strategies to optimise antibiotic prescribing? A pragmatic cluster-randomised controlled trial in primary care. *Qual Saf Health Care*. 2007;16(3):197-202.
48. van Eijk M, E., Avorn J, Porsius AJ, de Boer. Reducing prescribing of highly anticholinergic antidepressants for elderly people: randomised trial of group versus individual academic detailing. *BMJ*. 2001;322:654-657.

49. Verstappen WHJM, van der Weijden T, Sijbrandij J, Smeele I, Hermsen J, et al. Effect of a practice-based strategy on test ordering performance of primary care physicians: a randomized trial. *JAMA*. 2003;289(18):2407-2412.
50. Voorn VMA, Marang-van de Mheen PJ, van der Hout A, Hofstede SN, So-Osman C, et al. The effectiveness of a de-implementation strategy to reduce low-value blood management techniques in primary hip and knee arthroplasty: a pragmatic cluster-randomized controlled trial. *Implement Sci*. 2017;12(1):72.
51. Wei X, Zhang Z, Walley JD, Hicks JP, Zeng J, et al. Effect of a training and educational intervention for physicians and caregivers on antibiotic prescribing for upper respiratory tract infections in children at primary care facilities in rural China: a cluster-randomised controlled trial. *Lancet Glob Health*. 2017;5(12):e1258-e1267.
52. Welschen I, Kuyvenhoven Marijke M, Hoes Arno W, Verheij Theo JM. Effectiveness of a multiple intervention to reduce antibiotic prescribing for respiratory tract symptoms in primary care: randomised controlled trial. *BMJ*. 2004;329(7463):431.



# Chapter 8

## Effectiveness of strategies to reduce low-value medical tests in primary care: a systematic review

Toshihiko Takada*
Pauline Heus*
Sander van Doorn
Christiana A Naaktgeboren
Jan-Willem Weenink
Simone A van Dulmen
Lotty Hooft

*Authors contributed equally

*Submitted*

## Abstract

### Purpose

Overuse of medical tests in primary care is a recognized problem, however it is unclear how to reduce it. The aim of this study was to identify which strategies are effective in reducing the use of low-value medical tests in primary care settings.

### Methods

We searched MEDLINE, Embase and Rx for Change databases (January 1990 – November 2019) for randomized controlled trials (RCTs) evaluating strategies to reduce the use of low-value medical tests in primary care settings. Two reviewers selected eligible articles, extracted data, and assessed risk of bias.

### Results

Among included 16 RCTs, 11 studies reported a statistically significant reduction in the use of low-value medical tests. The median relative reduction was 17% (interquartile range 12 – 24). Strategies containing reminders or audit/feedback showed larger improvement than those without these components (22 vs. 14%, and 21 vs. 11%, respectively) and patient-targeted strategies showed larger reduction than those not targeted at patients (51 vs. 17%). Very few studies investigated sustainability of the effects, adverse events, cost-effectiveness, and patient-reported outcomes related to reducing low-value tests.

### Conclusions

This review suggests that it is possible to reduce the use of low-value medical tests in primary care, especially by using multiple components including reminders, audit/feedback, and patient-targeted interventions. Still, to widely implement these strategies in primary care settings, future studies need to investigate not only effectiveness, but also address adverse events, cost-effectiveness, and patient-reported outcomes.

## Background

In primary care settings, the use of medical tests is increasing.[1] However, a certain proportion of these tests is of low-value, providing no benefit to patients or even causing harm.[2,3]

Although primary care physicians are aware that they overuse medical tests,[4] there are some specific underlying mechanisms for this problem in primary care settings. First, as the pre-test probability of a serious disease is low and symptoms are overlapping between conditions, primary care physicians have to deal with greater diagnostic uncertainty than physicians in secondary and tertiary care settings.[5,6] Second, primary care plays a major role in delivering screening and monitoring (e.g., various types of cancers and lifestyle diseases). When tests that were once considered effective have been found to be ineffective, primary care physicians are expected to discontinue them (e.g., routine screening mammography in average risk women aged 40-49).[7] However, it is not easy to keep up with the emerging evidence in the broad field of medicine, in which primary care physicians are involved. Also, it has been reported that clinical guidelines have limited effect on physicians' practice.[8] Particularly, de-implementation (reducing the use of low-value care) of existing practice is sometimes more difficult than implementing new practices.[9]

In recent years, awareness of low-value care has increased and various initiatives to address the topic have been lauched.[10] Although several systematic reviews about interventions to reduce low-value care have been undertaken,[11-14] none of them specifically focused on reducing medical tests in primary care settings. Because of the distinctive challenges described above, a review of existing knowledge on this topic is necessary. Accordingly, the aim of this study was to identify which strategies are effective in reducing the use of low-value medical tests in primary care settings.



## Methods

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[15]

### Data Sources and Searches

This review is part of a larger project on de-implementation for which studies evaluating strategies to reduce low-value care were identified regardless of type of care (medical test or treatment), setting, or study design. An information specialist conducted a literature search using MEDLINE, Embase and Rx for Change databases on November 12th, 2019. The search strategy included synonyms for de-implementation

and low-value care (Supplemental Appendix). In addition, websites of healthcare quality improvement organizations were searched. We also used reference lists of all included studies and identified reviews on this topic as an additional source.

**Study Selection**

We included randomized controlled trials (RCTs) published in English, German, French, or Dutch after 1990, which evaluated the effectiveness of a strategy for reducing low-value medical tests in primary care settings. For protocols and conference abstracts, we checked whether the study had been published as a full text. Studies on guideline adherence were only included when the aim of the study was explicitly stated as reducing low-value healthcare practices. Pairs of authors independently screened titles and abstracts, and subsequently full texts of potentially eligible publications (TT, PH, CN, JWW, and SvDulmen). In the case of disagreement, the two authors discussed, and consulted a third author when necessary.

**Data Extraction and Critical Appraisal**

One of the authors (TT, PH, CN, JWW, SvDulmen, and LH) extracted data, which was checked by a second author. To ensure consistency between the reviewers, we used a structured, pilot-tested electronic data extraction form that included study characteristics (study design, the type of medical tests being de-implemented [e.g., laboratory/imaging/physiological], the role of tests [e.g., diagnosis/screening/staging/monitoring], targets and components of the de-implementation strategy) and outcomes. We classified target levels of a de-implementation strategy into four levels: provider, patient, organization, and healthcare system.[16] We divided the components of de-implementation strategies into nine categories according to the taxonomy provided by the Cochrane Effective Practice and Organisation of Care (EPOC) Group (Supplemental Appendix).[17,18] The primary outcome was the effect of strategies to reduce the use of low-value medical tests or the total number of tests. The secondary outcomes were adverse events due to unperformed medical tests (e.g., delay in diagnosis, referral, and treatment, and increased complication and mortality), other medical resource use (e.g., other medical tests, admission and visits to primary care/emergency room), cost-effectiveness of de-implementation strategies, and patient-reported outcomes (e.g., quality of life or patient satisfaction).

Two authors independently assessed risk of bias using the Cochrane Risk of bias tool (TT, PH, CN, JWW, SvDulmen, and LH).[19] In addition to the seven domains of this tool, we assessed three specific issues for cluster randomized trials: recruitment bias, unit of analysis error, and concern regarding baseline imbalances.[20-23]

### Analyses

The eligible studies reported the incidence of low-value medical tests in different ways (e.g., only incidence after intervention, difference between baseline and post-intervention, incidence per arm/practice/physician/visits/patients). To compare the effect of de-implementation strategies across the studies, we calculated the relative reduction in the use of the low-value tests (difference of the incidence between baseline and post-intervention divided by the incidence at baseline). Effectiveness of a strategy was defined as the difference between relative reductions in the intervention and control arms. The studies in which de-implementation strategies were directly compared with each other were reported separately to evaluate relative effectiveness of strategies. When a study investigated the effect of a strategy on several low-value tests, we selected the data of the low-value test with the median relative reduction as a representative of the study. In studies which compared several strategies, we selected the strategy including the most interventions or addressing the most targets. When there was only information about the total number of tests (without specifying if these were appropriate or inappropriate), we selected the relative reduction of total volume. In addition to the analysis of the effect of strategies at short term, we also assessed the sustainability of effects.

We explored factors potentially affecting the effect of strategies: type of medical tests (laboratory/imaging/physiological tests), role of tests (diagnostic/screening/staging/monitoring), number of intervention components, number of targets, outcome measured (total number of tests or actual low-value tests), overall risk of bias in the included studies, and the targets and components of the intervention. We defined studies with low overall risk of bias as satisfying all of the following criteria: 1) an adequate random sequence generation, 2) low risk of bias for all three domains related to cluster randomized designs, if applicable, and 3) not rated as high risk of bias due to unconcealed allocation, detection bias, attrition bias, or reporting bias, with unclear risk of bias for a maximum of two domains. We used Microsoft Office Excel version 16.16.9 for the data extraction form, R statistical software (version 3.6.0; R foundation for Statistical Computing, www.R-project.org) for summarizing the results, and Review Manager software version 5.3 for generating the risk of bias figures.

## Results

### Search results

Search results and the process of literature selection is shown in Figure 1. We identified 4590 records through the search. After excluding 3654 articles based on title and

abstract screening, we conducted a full text assessment for the remaining 936 articles, of which 16 were eligible for inclusion.[24-41]

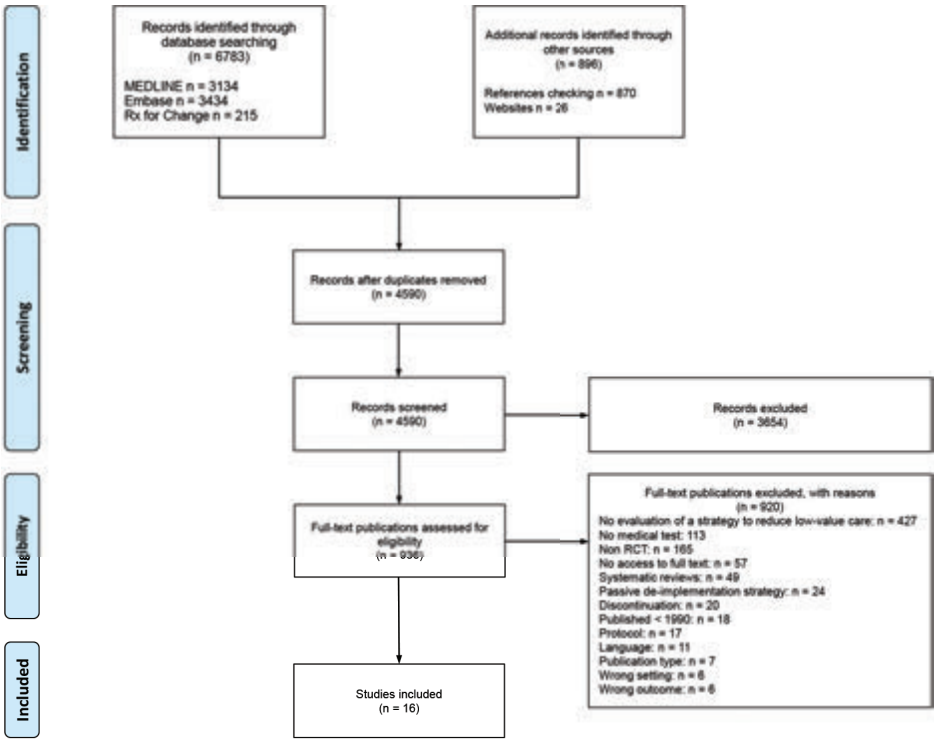



**Figure 1. Process of the literature selection**

### Characteristics of included studies

One-third of the studies were conducted in the United Kingdom (n=6; 38%) (Table 1). Fifteen (94%) were cluster-randomized design trials. Over half of the studies (n=9; 56%) specified the indications of the tests to be de-implemented, among which low back pain was the most common (n=4; 25%). The types of medical tests aimed for de-implementation were laboratory tests (n=7; 44%), imaging tests (n=13; 81%), and physiological tests (n=3; 19%). Four studies (25%) addressed multiple types of tests. Twelve studies (75%) specified the role of tests: diagnostics (n=12; 75%), screening (n=7; 44%), and monitoring (n=6; 38%) with some overlapping, while no study focused on tests for staging.

**Table 1. Summary of the characteristics of included studies**

| | Included studies n=16 |
|---|---|
| **Country** | |
| The United Kingdom | 6 (38) |
| The United States | 4 (25) |
| The Netherlands | 3 (19) |
| Australia | 2 (12) |
| Norway | 1 (6) |
| **Study Design** | |
| Cluster RCT | 15 (94) |
| RCT | 1 (6) |
| **Setting** | |
| Single center | 2 (12) |
| Multi-center | 14 (88) |
| **Indication for medical tests*** | |
| Specified | 9 (56) |
| Low back pain | 4 (25) |
| Others† | 5 (31) |
| Not specified | 7 (44) |
| **Type of Medical tests*** | |
| Laboratory tests | 7 (44) |
| Imaging | 13 (81) |
| Physiology | 3 (19) |

RCT = randomized controlled trial
Detailed information of each study is available in the Supplemental Appendix.
* Some studies were applicable to more than one category.
† Many other conditions were addressed, most of them were evaluated in only one of the included studies. The detailed information about each condition is shown in the Supplemental Appendix.



### De-implementation strategies

De-implementation strategies in the included studies were classified by their target and the number of interventions (single/combination of two or more) (Table 2). All six studies with a single target and a single intervention were aimed at healthcare providers. Among them, educational materials and reminders were most frequently used (33%). Similarly, healthcare providers were targeted in all seven studies having a single target and using a combination of interventions. Educational materials and audit/feedback were the most frequently used strategies (86% for both). Among three studies addressing multiple targets with a combination of interventions, all targeted the healthcare provider. Two (67%) studies additionally targeted patients and one study

additionally targeted the organizational context and healthcare system. Educational meetings and materials were used in all three studies. Detailed information of each study is available in the Supplemental Appendix.

**Table 2. De-implementation strategies by the number of its intervention components and targets**

| Intervention | Single target, single intervention N=6 (%) | Single target, combination of interventions N=7 (%) | Multiple targets, combination of interventions N=3 (%) | All N=16 (%) |
|---|---|---|---|---|
| *Targeted at provider* | 6 (100) | 7 (100) | 3 (100) | 16 (100) |
| Educational meetings | 1 (17) | 4 (57) | 3 (100) | 8 (50) |
| Distribution of educational material | 2 (33) | 6 (86) | 3 (100) | 11 (69) |
| Reminders | 2 (33) | 2 (29) | 1 (33) | 5 (31) |
| Audit/feedback | 1 (17) | 6 (86) | 1 (33) | |
| Financial interventions | | | | 0 (0) |
| *Targeted at patient* | 0 (0) | 0 (0) | 2 (67) | 2 (13) |
| *Targeted at organizational context* | 0 (0) | 0 (0) | 1 (33) | 1 (6) |
| Organizational interventions | 0 (0) | 0 (0) | 1 (33) | 1 (6) |
| Structural interventions | - | - | - | 0 (0) |
| *Targeted at healthcare system* | 0 (0) | 0 (0) | 1 (33) | 1 (6) |
| Regulatory interventions | - | - | - | 0 (0) |
| Financial interventions | 0 (0) | 0 (0) | 1 (33) | 1 (6) |

**Risk of bias**

The results of the assessment of risk of bias are shown in the Supplemental Appendix. In the domain of allocation concealment, seven studies (44%) were rated as low-risk of bias, while nine studies (56%) did not give sufficient information. Since blinding of participants was difficult due to the nature of the intervention, most studies (69%) were rated as high-risk of bias in this item. Four studies (25%) satisfied the criteria of overall low risk of bias.

**Effectiveness of de-implementation**

Eleven studies (69%) reported that their intervention showed statistically significant reduction. Ten studies (63%) reported the necessary information to calculate relative reductions of the incidence of the low-value tests. The results of the six studies without information for calculation of relative reduction are summarized in the Supplemental Appendix.

*Comparison of de-implementation to usual care*

The median of relative reductions in the use of low-value tests was 17% (interquartile range [IQR]: 12 – 24) (Table 3). Strategies with multiple targets and a combination of interventions tended to be more effective than those with a single target. Strategies using reminders and audit/feedback showed a larger reduction than those without these components (22% [IQR 17 - 31] vs. 14% [IQR 12 - 20], and 21% [IQR 14 - 31] vs. 11% [IQR 10 -12], respectively). Studies targeted at patients showed a larger reduction of low-value tests than those not targeted at patients (51% [IQR 30 - 72] vs. 17% [IQR 12 - 23]).

**Table 3. Differences in relative reductions between de-implementation and usual care**

| | N | Median (IQR) |
|---|---|---|
| **All** | 10 | 17 (12 - 24) |
| **Type of medical tests** | | |
| Laboratory tests | 3 | 22 (18 - 24) |
| Imaging | 5 | 13 (10 - 50) |
| Laboratory test, imaging, and physiology | 2 | 15 (13 - 18) |
| **Role of medical tests** | | |
| Diagnosis | 3 | 22 (16 - 57) |
| Diagnosis, screening, and monitoring | 4 | 14 (12 - 17) |
| Unspecified | 3 | 20 (15 - 35) |
| **Number of intervention components** | | |
| 1 | 3 | 13 (12 - 16) |
| 2 | 2 | 12 (11 - 13) |
| 3 or more | 5 | 25 (22 - 50) |
| **Single or multiple targets and interventions** | | |
| Single target, single intervention | 3 | 13 (12 - 16) |
| Single target, combination of interventions | 5 | 22 (15 - 25) |
| Multiple targets, combination of interventions | 2 | 51 (30 - 72) |
| **Outcome measured** | | |
| Low-value care | 3 | 22 (17 - 57) |
| Total volume of care | 7 | 15 (11 - 22) |
| **Bias** | | |
| Low | 2 | 18 (17 - 20) |
| High | 8 | 16 (11 - 31) |
| **Intervention categories and targets** | | |
| Targeted at provider | 10 | 17 (12 - 24) |
| Educational component (either meetings or materials, or both) | | |
| Yes | 8 | 18 (12 - 31) |
| No | 2 | 15 (13 - 18) |
| Reminders | | |
| Yes | 4 | 22 (17 - 31) |

| | | |
|---|---|---|
| No | 6 | 14 (12 - 20) |
| Audit/feedback | | |
| Yes | 8 | 21 (14 - 31) |
| No | 2 | 11 (10 - 12) |
| Targeted at patient | 2 | 51 (30 - 72) |
| Not targeted at patient | 8 | 17 (12 - 23) |

IQR = interquartile range

*Direct comparison of de-implementation strategies*

In three studies, a direct comparison of de-implementation strategies was reported.[26,33,36] In one study, reminders were more effective than audit/feedback in de-implementation of imaging studies (41% vs. 29% for lumbar radiograph, and 33% vs. 15% for knee radiograph, respectively).[26] However, another study showed an opposite trend that reminders were less effective than audit/feedback in de-implementation of laboratory tests (15% vs. 27%, respectively).[33] In the other study, a computer-based decision support system based on the guidelines reduced the number of laboratory tests by 20% compared to a system based on a reduced list of medical tests.[36]

**Sustainability of effect**

Three studies evaluated the sustainability of the effect of the strategy.[34,40,41] One of them did not report results.[41] The two others reported that the effect of the strategy was not sustainable (5 -12 months after intervention) despite an initially observed significant effect.[34,40]

**Secondary outcomes**

One study reported adverse events due to unperformed tests and found no increase in the number of hospitalizations, emergency room visits, and outpatient visits.[34]

One study assessed the cost-effectiveness of de-implementation strategies. In the comparison between an original multifaceted strategy (combining written feedback, group education, and distribution of guidelines) and only feedback, the multifaceted strategy was more effective in cost reduction than only feedback. However, the cost for the strategy surpassed the reduced cost.[39]

Two studies measured patient satisfaction. In one study, the intervention was designed to enhance primary care physicians' patient-centeredness and skills in handling patient requests for low-value diagnostic tests. Patients in the intervention group were more satisfied than in those in the control group.[27] The other study stated in the method section to measure patient satisfaction, however, no results were reported.[32]

## Discussion

Among 16 RCTs investigating the effect of strategies to reduce low-value medical tests in primary care, 11 studies (69%) reported a statistically significant reduction. The median relative reduction was 17% (IQR 12 - 24). Addressing multiple targets and using a combination of interventions tended to increase effectiveness. Strategies containing reminders or audit/feedback showed larger improvement than those without these components (22 vs. 14%, and 21 vs. 11%, respectively) and patient-targeted strategies showed a larger reduction of low-value tests than those not targeted at patients (51 vs. 17%).

Our findings corroborate the results of the existing systematic reviews about strategies to promote the appropriate use of medical tests, which included mainly observational studies without a control group. Some of these reviews showed that interventions to reduce laboratory test utilization are generally successful.[11,14,42] However, they focused only on laboratory tests and the setting of two reviews was solely[11] or mainly secondary/tertiary care.[14] Another review showed that multicomponent interventions were more effective than single component interventions in increasing appropriate use of diagnostic tests by physicians in various settings.[43] However, this review included not only studies which aimed at reducing low-value tests, but it also addressed studies which aimed at promoting underused tests. While there have been several reviews on quality improvement in primary care,[44] to the best of our knowledge, this is the first review to evaluate the effect of strategies to reduce low-value medical tests in primary care.



In line with findings of a review that evaluated the effect of de-implementation strategies with no restriction on types of low-value care (medical tests or treatment) and settings,[13] we showed that strategies targeting not only providers but also patients are more effective. While physicians may order low-value tests due to their diagnostic uncertainty or misconceptions of the value of tests, patients may frequently request those tests themselves. It has been reported that such patients are usually anxious and require reassurance.[45] Although physicians sometimes rationalize the use of low-value tests to reassure patients, these tests hardly help to decrease patients' anxiety.[46] To improve patients' understanding of low-value tests, our results suggest that it is of added value to include patient educational components as a part of de-implementation strategies.

While the effect of de-implementation strategies on usage of low-value care has been extensively evaluated, there is little evidence on potential negative consequences of

these strategies. One of the reasons for this might be that negative effects are rare, and need a large sample size and long follow-up to be evaluated. Furthermore, it might not be easy to track patients who are referred to or visit other clinics or hospitals. Since fear for juridical claims is one of the reasons for physicians to order tests,[47] it is necessary to assure that low-value tests can be omitted without adverse events (e.g., delay in diagnosis, referral, and treatment, and increased complication and mortality). Furthermore, sustainability and cost-effectiveness are crucial considerations for introducing de-implementation strategies at a larger scale. Nevertheless, there were only a few studies that evaluated these outcomes. Also, patient satisfaction is an important outcome in clinical practice, that can be impaired by declining medical tests requested by patients.[48] More studies assessing these consequences of interventions on a long term are needed before the spread of de-implementation strategies for low-value tests is promoted.

Our study has several limitations. First, for six studies we could not calculate the relative reduction of the use of low-value tests, as they lacked necessary information. This has also been encountered in other reviews.[11,14] To promote the integration of evidence, recommendations about appropriate outcome measures for de-implementation are required. Second, in the analyses about factors related to the effect of strategies, there was a very small number of studies in some categories, leading to less precision. Finally, there was substantial heterogeneity among the included studies in terms of type and role of medical tests, components and targets of intervention. As a result, it was difficult to disentangle the effect of each of these factors.

In conclusion, despite the specific challenges in primary care settings, this review suggests that it is possible to reduce low-value medical tests in primary care, especially by combining multiple intervention components, including reminders and audit/feedback, and targeting patients. Still, to widely implement these strategies in primary care settings, future studies need to investigate sustainability of the effect, adverse events, cost-effectiveness, and patient-reported outcomes as consequences of de-implementation of low-value medical tests.

## Acknowledgements

We thank René Spijker for developing and performing the searches, and Eva Verkerk and Tijn Kool for their contribution to the selection of articles and data extraction.

## References


1. O'Sullivan JW, Stevens S, Hobbs FDR, et al. Temporal trends in use of tests in UK primary care, 2000-15: retrospective analysis of 250 million tests. *BMJ*. 2018;363:k4666.
2. de Vries EF, Struijs JN, Heijink R, Hendrikx RJ, Baan CA. Are low-value care measures up to the task? A systematic review of the literature. *BMC Health Serv Res*. 2016;16:405.
3. O'Sullivan JW, Albasri A, Nicholson BD, et al. Overtesting and undertesting in primary care: a systematic review and meta-analysis. *BMJ Open*. 2018;8:e018557.
4. Sirovich BE, Woloshin S, Schwartz LM. Too Little? Too Much? Primary care physicians' views on US health care: a brief report. *Arch Intern Med*. 2011;171:1582-1585.
5. Okkes IM, Oskam SK, Lamberts H. The probability of specific diagnoses for patients presenting with common symptoms to Dutch family physicians. *J Fam Pract*. 2002;51:31-36.
6. McCowan C, Fahey T. Diagnosis and diagnostic testing in primary care. *Br J Gen Pract*. 2006;56:323-324.
7. Siu AL, Force USPST. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2016;164:279-296.
8. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA*. 1999;282:1458-1465.
9. Roman BR, Asch DA. Faded promises: the challenge of deadopting low-value care. *Ann Intern Med*. 2014;161:149-150.
10. Choosing Wisely. https://www.choosingwisely.org. Accessed February 3rd, 2020.
11. Bindraban RS, Ten Berg MJ, Naaktgeboren CA, Kramer MHH, Van Solinge WW, Nanayakkara PWB. Reducing Test Utilization in Hospital Settings: A Narrative Review. *Ann Lab Med*. 2018;38:402-412.
12. Chambers JD, Salem MN, D'Cruz BN, Subedi P, Kamal-Bahl SJ, Neumann PJ. A Review of Empirical Analyses of Disinvestment Initiatives. *Value Health*. 2017;20:909-918.
13. Colla CH, Mainor AJ, Hargreaves C, Sequist T, Morden N. Interventions Aimed at Reducing Use of Low-Value Health Services: A Systematic Review. *Med Care Res Rev*. 2017;74:507-550.
14. Kobewka DM, Ronksley PE, McKay JA, Forster AJ, van Walraven C. Influence of educational, audit and feedback, system based, and incentive and penalty interventions to reduce laboratory test utilization: a systematic review. *Clin Chem Lab Med*. 2015;53:157-183.

15. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med*. 2009;151:W65-94.
16. Grol RW, M.; Eccles, MP; Davis, D. *Improving patient care: the implementation of change in health care.* John Wiley & Sons; 2013.
17. Effective Practice and Organisation of Care (EPOC); EPOC Taxonomy. https://epoccochraneorg/epoc-taxonomy. Accessed February 3rd, 2020.
18. Effective Practice and Organisation of Care (EPOC). *The EPOC taxonomy of health systems interventions. EPOC Resources for review authors*. Oslo, Norwegian Knowledge Centre for the Health Services; 2016.
19. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928.
20. Bolzern J, Mnyama N, Bosanquet K, Torgerson DJ. A review of cluster randomized trials found statistical evidence of selection bias. *J Clin Epidemiol*. 2018;99:106-112.
21. Brierley G, Brabyn S, Torgerson D, Watson J. Bias in recruitment to cluster randomized trials: a review of recent publications. *J Eval Clin Pract*. 2012;18:878-886.
22. Higgins JPT. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester: John Wiley & Sons; 2019.
23. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ*. 2003;327:785-789.
24. Bearcroft PW, Small JH, Flower CD. Chest radiography guidelines for general practitioners: a practical approach. *Clin Radiol*. 1994;49:56-58.
25. Dey P, Simpson CW, Collins SI, et al. Implementation of RCGP guidelines for acute low back pain: a cluster randomised controlled trial. *Br J Gen Pract*. 2004;54:33-37.
26. Eccles M, Steen N, Grimshaw J, et al. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. *Lancet*. 2001;357:1406-1409.
27. Fenton JJ, Kravitz RL, Jerant A, et al. Promoting Patient-Centered Counseling to Reduce Use of Low-Value Diagnostic Tests: A Randomized Clinical Trial. *JAMA Intern Med*. 2016;176:191-197.
28. Flottorp S, Oxman AD, Havelsrud K, Treweek S, Herrin J. Cluster randomised controlled trial of tailored interventions to improve the management of urinary tract infections in women and sore throat. *BMJ*. 2002;325:367.
29. French SD, McKenzie JE, O'Connor DA, et al. Evaluation of a theory-informed implementation intervention for the management of acute low back pain in general medical practice: the IMPLEMENT cluster randomised trial. *PLoS One*. 2013;8:e65471.

30. Kerry S, Oakeshott P, Dundas D, Williams J. Influence of postal distribution of the Royal College of Radiologists' guidelines, together with feedback on radiological referral rates, on X-ray referrals from general practice: a randomized controlled trial. *Fam Pract.* 2000;17:46-52.
31. Oakeshott P, Kerry SM, Williams JE. Randomized controlled trial of the effect of the Royal College of Radiologists' guidelines on general practitioners' referrals for radiographic examination. *Br J Gen Pract.* 1994;44:197-200.
32. Schectman JM, Schroth WS, Verme D, Voss JD. Randomized controlled trial of education and feedback for implementation of guidelines for acute low back pain. *J Gen Intern Med.* 2003;18:773-780.
33. Thomas RE, Croal BL, Ramsay C, Eccles M, Grimshaw J. Effect of enhanced feedback and brief educational reminder messages on laboratory test requesting in primary care: a cluster randomised trial. *Lancet.* 2006;367:1990-1996.
34. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. *N Engl J Med.* 1990;322:1499-1504.
35. Trietsch J, van Steenkiste B, Grol R, et al. Effect of audit and feedback with peer review on general practitioners' prescribing and test ordering performance: a cluster-randomized controlled trial. *BMC Fam Pract.* 2017;18:53.
36. van Wijk MA, van der Lei J, Mosseveld M, Bohnen AM, van Bemmel JH. Assessment of decision support for blood test ordering in primary care. a randomized trial. *Ann Intern Med.* 2001;134:274-281.
37. Verstappen WH, van der Weijden T, Dubois WI, et al. Improving test ordering in primary care: the added value of a small-group quality improvement strategy compared with classic feedback only. *Ann Fam Med.* 2004;2:569-575.
38. Verstappen WH, van der Weijden T, Sijbrandij J, et al. Effect of a practice-based strategy on test ordering performance of primary care physicians: a randomized trial. *JAMA.* 2003;289:2407-2412.
39. Verstappen WH, van Merode F, Grimshaw J, Dubois WI, Grol RP, van der Weijden T. Comparing cost effects of two quality strategies to improve test ordering in primary care: a randomized trial. *Int J Qual Health Care.* 2004;16:391-398.
40. Weller D, May F, Rowett D, et al. Promoting better use of the PSA test in general practice: randomized controlled trial of educational strategies based on outreach visits and mailout. *Fam Pract.* 2003;20:655-661.
41. Winkens RA, Pop P, Bugter-Maessen AM, et al. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *Lancet.* 1995;345:498-502.
42. Cadogan SL, Browne JP, Bradley CP, Cahill MR. The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review. *Implement Sci.* 2015;10:167.

43. Solomon DH, Hashimoto H, Daltroy L, Liang MH. Techniques to improve physicians' use of diagnostic tests: a new conceptual framework. *JAMA*. 1998;280:2020-2027.
44. Irwin R, Stokes T, Marshall T. Practice-level quality improvement interventions in primary care: a review of systematic reviews. *Prim Health Care Res Dev*. 2015;16:556-577.
45. van Bokhoven MA, Pleunis-van Empel MC, Koch H, Grol RP, Dinant GJ, van der Weijden T. Why do patients want to have their blood tested? A qualitative study of patient expectations in general practice. *BMC Fam Pract*. 2006;7:75.
46. Rolfe A, Burton C. Reassurance after diagnostic testing with a low pretest probability of serious disease: systematic review and meta-analysis. *JAMA Intern Med*. 2013;173:407-416.
47. van der Weijden T, van Bokhoven MA, Dinant GJ, van Hasselt CM, Grol RP. Understanding laboratory testing in diagnostic uncertainty: a qualitative study in general practice. *Br J Gen Pract*. 2002;52:974-980.
48. Kravitz RL, Bell RA, Azari R, Krupat E, Kelly-Reif S, Thom D. Request fulfillment in office practice: antecedents and relationship to outcomes. *Med Care*. 2002;40:38-51.

## Supplemental Appendix

### Section 1: Search strategy

Conducted by René Spijker on November 12, 2019

For data sources and search terms see Supplemental appendix of Chapter 6 (page 180).

### Section 2: Classification of components of de-implementation strategies based on the taxonomy provided by the Cochrane Effective Practice and Organisation of Care (EPOC) Group[1,2]

See Supplementary material of Chapter 7: Classification of interventions (page 220).

**Section 3: Detailed information of each study**

| Study | Country | Study design | Unit of randomization | Setting | Target condition |
|---|---|---|---|---|---|
| Bearcroft 1994[3] | UK | Cluster RCT | Practice | Multi-center | Five indications for chest radiography |
| Dey 2004[4] | UK | Cluster RCT | Health center | Multi-center | Acute low back pain |
| Eccles 2001[5] | UK | Cluster RCT | Practice | Multi-center | No specific target condition |
| Fenton 2016[6] | USA | Cluster RCT | Physician | Single center | Low back pain/ postmenopausal women at low risk of osteoporosis/ Neuroimaging for recent-onset headache |
| Flottorp 2002[7] | Norway | Cluster RCT | Practice | Multi-center | Urinary tract infection and sore throat |
| French 2013[8] | Australia | Cluster RCT | Practice | Multi-center | Acute low back pain |
| Kerry 2000[9] | UK | Cluster RCT | Practice | Multi-center | No specific condition |
| Oakeshott 1994[10] | UK | Cluster RCT | Practice | Multi-center | No specific condition |
| Schectman 2003[11] | USA | Cluster RCT | Practice | Multi-center | Acute low back pain |
| Thomas 2006[12] | UK | Cluster RCT | Practice | Multi-center | No specific target condition |
| Tierney 1990[13] | USA | Cluster RCT | Physician | Single center | No specific target condition |

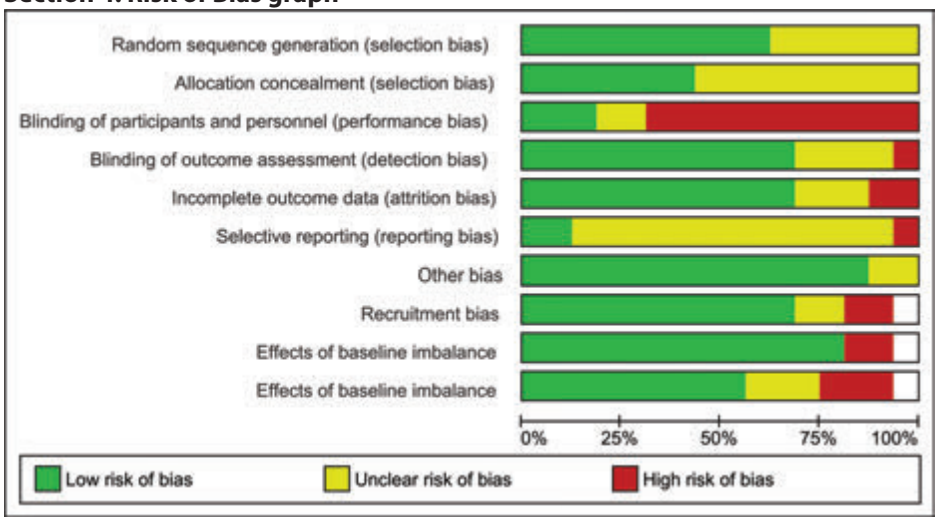
| Aim of medical tests | Medical tests | Target of intervention | Intervention | Overall risk of bias |
|---|---|---|---|---|
| Unspecified | Chest X ray | Provider | Educational materials | High |
| Diagnosis | Lumbar spine X ray | Provider and organization | Educational meetings, educational materials, organizational interventions | High |
| Unspecified | Lumbar spine and knee X ray | Provider | Educational materials, reminders, audit/ feedback | High |
| Diagnosis, screening | MRI for low back pain, DXA for postmenopausal women, and neuroimaging for recent onset headache | Provider | Educational meetings | High |
| Diagnosis | Laboratory tests (for evaluation of sore throat and urinary tract infection | Provider, patient and system | Educational meetings, educational materials, reminders, patient-targeted interventions, financial incentives | High |
| Diagnosis | X ray and CT for evaluation of acute low back pain | Provider | Educational meetings, educational materials | High |
| Unspecified | X ray of chest or limbs or spine | Provider | Educational materials, audit/feedback | High |
| Unspecified | X ray of chest or limbs or spine | Provider | Educational materials | High |
| Diagnosis | X ray, CT, and MRI for evaluation of acute low back pain | Provider and patient | Educational meetings, educational materials, audit/feedback, patient-targeted interventions | High |
| Unspecified | Laboratory tests (AAS, CA125, CEA, Ferritin, FSH, HPS, IgE, TSH, Vitamin B12) | Provider | Educational materials, reminders, audit/ feedback | High |
| Unspecified | All laboratory tests and imaging performed by the clinical laboratory/ radiology/nuclear medicine/diagnostic cardiology department | Provider | Reminders | High |

| | | | | | |
|---|---|---|---|---|---|
| Trietsch 2017[14] | Netherlands | Cluster RCT | Local quality improvement collaboratives | Multi-center | Anemia, rheumatic complaints, prostate complaints, Chlamydia infections, thyroid dysfunction, and perimenopausal conditions |
| Van Vijk 2001[15] | Netherlands | Cluster RCT | Practice | Multi-center | No specific target condition |
| Verstappen 2003/2004[16-18] | Netherlands | Cluster RCT | Physician group | Multi-center | Cardiovascular/ hypertension, upper/ lower abdominal complaints, COPD/asthma, general complaints, and degenerative joint complaints |
| Weller 2003[19] | Australia | Cluster RCT | Practice | Multi-center | Prostate cancer |
| Winkens 1995[20] | Netherlands | RCT | Physician | Multi-center | No specific target condition |

UK = United Kingdom; USA = United States of America; RCT = randomized controlled trial; MRI = magnetic resonance imaging; DXA = dual energy X-ray absorptiometry; CT = computed tomography; AAS = autoantibody screening; CA125 = carbohydrate antigen-125; CEA = carcino-embryonic antigen; FSH = follicle stimulating hormone; HPS = Helicobacter pylori serology; TSH = thyroid stimulating hormone; COPD = chronic obstructive pulmonary disease; PSA = prostate specific antigen

| | | | | |
|---|---|---|---|---|
| Unspecified | Various laboratory tests in evaluation of anemia, rheumatic complaints, prostate complaints, Chlamydia infections, thyroid dysfunction, and perimenopausal conditions | Provider | Educational meetings, audit/feedback | Low |
| Unspecified | All laboratory tests | Provider | Reminders | Low |
| Diagnosis, screening | Various tests in evaluation of cardiovascular/ hypertension, upper/ lower abdominal complaints, COPD/asthma, general complaints, and degenerative joint complaints | Provider | Educational meetings, educational materials, audit/feedback | Low |
| Unspecified | PSA | Provider | Educational meetings, educational materials, audit/feedback | Low |
| Unspecified | Electrography/Endoscopy/ Cervical smears/Allergy tests/Radiography/ Ultrasound | Provider | Audit/feedback | High |

**Section 4: Risk of Bias graph**



**Section 5: Risk of Bias summary**

**Section 6: Detailed information of studies without information for calculation of relative reduction**

| Study | Aim of medical tests | Target condition | Medical tests | Target of intervention | Intervention | Outcome description | Effect of intervention compared to the control group reported by authors |
|---|---|---|---|---|---|---|---|
| Bearcroft 1994[3] | Unspecified | Five indications for chest radiography | Chest X ray | Provider | Educational materials | Referrals to chest X ray which were contrary to the guidelines | Reduction by 30.5% (P = 0.016) |
| Dey 2004[4] | Diagnosis | Acute low back pain | Lumbar spine X ray | Provider and organization | Educational meetings, educational materials, organizational interventions | Referrals to lumbar spine X ray | Difference between the intervention and control groups were 1.4% (95% CI -4.1 - 6.8) |
| Fenton 2016[6] | Diagnosis, screening | Low back pain/ postmenopausal women at low risk of osteoporosis/ Neuroimaging for recent-onset headache | MRI for low back pain, DXA for postmenopausal women, and neuroimaging for recent onset headache | Provider | Educational meetings | Ordering of medical tests | Adjusted odds ratio of 1.07 (95% CI 0.49 - 2.32) |
| French 2013[8] | Diagnosis | Acute low back pain | X ray and CT for evaluation of acute low back pain | Provider | Educational meetings, educational materials | Referral to X ray or CT scan | Incidence rate ratio of 0.87 (95% CI 0.68 - 1.10) |

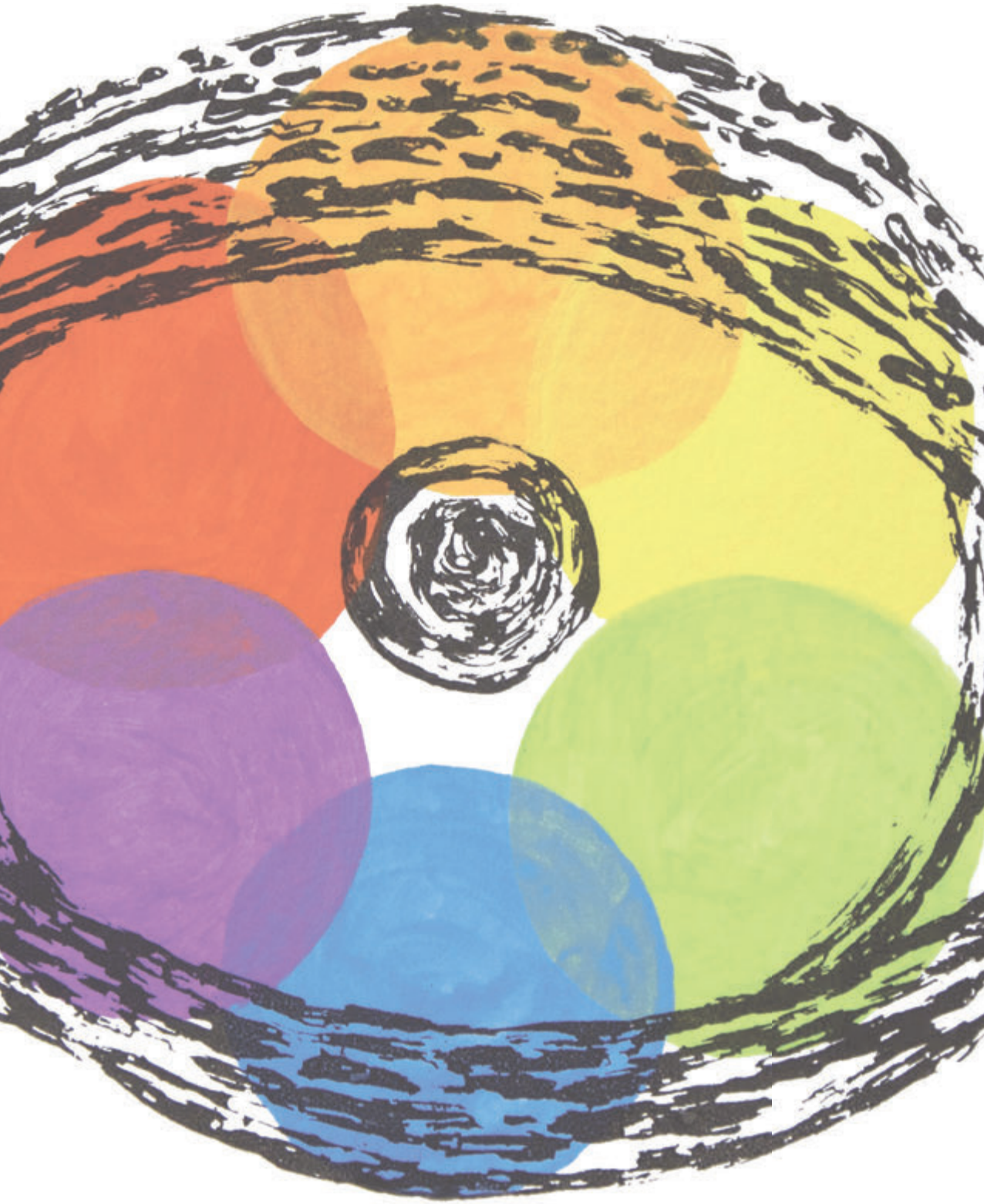
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| van Vijk 2001[15] | Unspecified | No specific target condition | All laboratory tests | Provider | Reminders (Comparison between decision support systems based on a reduced list of tests vs. the guidelines) | Number of laboratory tests per order form per practice | General practitioners who used the system based on guideline requested 20% fewer tests (mean ± SD, 5.5 ± 0.9 vs. 6.9 ± 1.6 tests, P = 0.003) |
| Weller 2003[19] | Unspecified | Prostate cancer | PSA | Provider | Educational meetings, educational materials, audit/feedback | PSA test ordering rates per 100 consultations | At 6 months after the intervention, PSA test ordering rates for educational outreach visits, posted materials, and the control group were 1.32, 1.97, and 1.78, respectively, with statistically significance. |

DXA = dual energy X-ray absorptiometry; CT = computed tomography; PSA = prostate specific antigen; CI = confidence interval

**References**


1. Effective Practice and Organisation of Care (EPOC); EPOC Taxonomy. https://epoccochraneorg/epoc-taxonomy. Accessed February 3rd, 2020.
2. Effective Practice and Organisation of Care (EPOC). *The EPOC taxonomy of health systems interventions. EPOC Resources for review authors*. Oslo, Norwegian Knowledge Centre for the Health Services; 2016.
3. Bearcroft PW, Small JH, Flower CD. Chest radiography guidelines for general practitioners: a practical approach. *Clin Radiol.* 1994;49:56-58.
4. Dey P, Simpson CW, Collins SI, et al. Implementation of RCGP guidelines for acute low back pain: a cluster randomised controlled trial. *Br J Gen Pract.* 2004;54:33-37.
5. Eccles M, Steen N, Grimshaw J, et al. Effect of audit and feedback, and reminder messages on primary-care radiology referrals: a randomised trial. *Lancet*. 2001;357:1406-1409.
6. Fenton JJ, Kravitz RL, Jerant A, et al. Promoting Patient-Centered Counseling to Reduce Use of Low-Value Diagnostic Tests: A Randomized Clinical Trial. *JAMA Intern Med*. 2016;176:191-197.
7. Flottorp S, Oxman AD, Havelsrud K, Treweek S, Herrin J. Cluster randomised controlled trial of tailored interventions to improve the management of urinary tract infections in women and sore throat. *BMJ*. 2002;325:367.
8. French SD, McKenzie JE, O'Connor DA, et al. Evaluation of a theory-informed implementation intervention for the management of acute low back pain in general medical practice: the IMPLEMENT cluster randomised trial. *PLoS One*. 2013;8:e65471.
9. Kerry S, Oakeshott P, Dundas D, Williams J. Influence of postal distribution of the Royal College of Radiologists' guidelines, together with feedback on radiological referral rates, on X-ray referrals from general practice: a randomized controlled trial. *Fam Pract*. 2000;17:46-52.
10. Oakeshott P, Kerry SM, Williams JE. Randomized controlled trial of the effect of the Royal College of Radiologists' guidelines on general practitioners' referrals for radiographic examination. *Br J Gen Pract*. 1994;44:197-200.
11. Schectman JM, Schroth WS, Verme D, Voss JD. Randomized controlled trial of education and feedback for implementation of guidelines for acute low back pain. *J Gen Intern Med*. 2003;18:773-780.
12. Thomas RE, Croal BL, Ramsay C, Eccles M, Grimshaw J. Effect of enhanced feedback and brief educational reminder messages on laboratory test requesting in primary care: a cluster randomised trial. *Lancet*. 2006;367:1990-1996.
13. Tierney WM, Miller ME, McDonald CJ. The effect on test ordering of informing physicians of the charges for outpatient diagnostic tests. *N Engl J Med*. 1990;322:1499-1504.

14. Trietsch J, van Steenkiste B, Grol R, et al. Effect of audit and feedback with peer review on general practitioners' prescribing and test ordering performance: a cluster-randomized controlled trial. *BMC Fam Pract.* 2017;18:53.
15. van Wijk MA, van der Lei J, Mosseveld M, Bohnen AM, van Bemmel JH. Assessment of decision support for blood test ordering in primary care. a randomized trial. *Ann Intern Med.* 2001;134:274-281.
16. Verstappen WH, van der Weijden T, Dubois WI, et al. Improving test ordering in primary care: the added value of a small-group quality improvement strategy compared with classic feedback only. *Ann Fam Med.* 2004;2:569-575.
17. Verstappen WH, van der Weijden T, Sijbrandij J, et al. Effect of a practice-based strategy on test ordering performance of primary care physicians: a randomized trial. *JAMA.* 2003;289:2407-2412.
18. Verstappen WH, van Merode F, Grimshaw J, Dubois WI, Grol RP, van der Weijden T. Comparing cost effects of two quality strategies to improve test ordering in primary care: a randomized trial. *Int J Qual Health Care.* 2004;16:391-398.
19. Weller D, May F, Rowett D, et al. Promoting better use of the PSA test in general practice: randomized controlled trial of educational strategies based on outreach visits and mailout. *Fam Pract.* 2003;20:655-661.
20. Winkens RA, Pop P, Bugter-Maessen AM, et al. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *Lancet.* 1995;345:498-502.



# Chapter 9

## General discussion

This thesis explored two important components of evidence-based medicine (EBM): the transparent and accurate reporting of research and the implementation of research findings into practice.

With regard to research reporting we focussed on the reporting of prediction model studies and the TRIPOD reporting guideline that was developed to enhance transparent and accurate reporting of diagnostic and prognostic prediction model studies. The following lessons can be learned from the first part of this thesis:

- The reporting of prediction model studies is poor: more than half of the items that are considered essential according to TRIPOD were not fully addressed in publications of multivariable prediction model studies. Especially the information with regard to title, abstract, statistical analysis methods, and results (i.e. model specifications and model performance) was often not detailed enough, which reduces the usability and generalisability of prediction models in both practice and further research (Chapter 2).
- A so-called adherence assessment form including guidance and scoring rules is essential to ensure consistency between guideline-adherence evaluations and facilitate adherence comparisons over time, as well as between different clinical fields. The TRIPOD adherence assessment form we developed should be used by anyone (e.g., researchers, reviewers, editors) evaluating the adherence of published prediction model studies to TRIPOD, to make these assessments comparable regardless of the type of prediction model study and clinical domain (Chapter 3).
- Following our findings in Chapter 2 of incomplete reporting of titles and abstracts, we developed a reporting checklist and corresponding guidance, which are applicable to journal and conference abstracts that describe the development, external validation, update or extension of a diagnostic or prognostic prediction model, regardless the clinical domain or statistical approach used (Chapter 4). Titles and abstracts are essential elements of a study report, facilitating identification as well as judgement of the relevance and importance of a study.
- Almost two thirds of medical journals endorse one or more reporting guidelines and TRIPOD was endorsed by 9%. Editors of medical journals suggested the following to overcome barriers to the use and endorsement of reporting guidelines: make adherence or use mandatory for authors and reviewers; education and dissemination of tools how to use the reporting guideline; and the use of software applications and automated tools for

identifying reporting guidelines and checking publications on guideline adherence (Chapter 5).

The second part of this thesis addressed implementation of research findings that recommends to abandon the routine use of a specific healthcare practice, so called de-implementation. The following lessons can be learned from the second part of this thesis:

- Barriers and facilitators to de-implementation of healthcare practices are for a large part related to the individual healthcare provider, and rather related to attitude, than to knowledge or behaviour (Chapter 6).
- Patient-provider interaction, the fear of consequences of providing incorrect care, and financial incentives are more important barriers to de-implementation than for implementation of specific healthcare practices (Chapter 6).
- Many healthcare de-implementation strategies achieve a considerable reduction of low-value care, especially those applying a multifaceted strategy. It seems worthwhile to consider audit & feedback and patient-directed interventions as components of a de-implementation strategy (Chapter 7). For reducing the use of low-value medical tests in primary care also reminders appeared to be a potential effective strategy component (Chapter 8).
- Details regarding sustainability of effect and impact of a de-implementation strategy on health outcomes are often not evaluated in de-implementation studies. This is, however, essential information for interpretation and application of findings with regard to rolling out successful de-implementation strategies at a larger scale (Chapters 7 and 8).

**Implications for practice and research**

In the following sections I will reflect upon the lessons learned of this thesis, thereby addressing ways to promote the use of reporting guidelines in general and TRIPOD in particular. In addition, I will discuss challenges regarding the design and evaluation of de-implementation strategies and illustrate this with a case study.

*Reporting of prediction model studies*

With the growing interest in personalized medicine, it is likely that the role of prediction models which can predict the probability of an individual having a certain outcome (diagnostic models) or developing a certain outcome (prognostic models) will become

increasingly important. Transparent and accurate reporting of these models is not only essential for the application in daily clinical decision making, but also when conducting evidence syntheses or reviews to summarize the rising number of available prediction models in a certain clinical context or domain. In addition, detailed reporting of the development of a prediction model is needed for the assessment of the model's predictive performance in new individuals (external validation). It is worrisome that our assessment of reporting revealed that especially essential items required for identification (title and abstract), external validation (model building procedures), and application in clinical practice (model specifications) of prediction models were among the least well reported items. Improved adequacy of reporting is thus urgently needed.

Reporting guidelines can positively contribute to the way biomedical studies are reported, although we know that reporting improvements in the biomedical literature are slow and modest.[1-3] To determine the impact of the TRIPOD statement, adherence to it should be measured and monitored over time. As we used a set of prediction model studies published before the introduction of the TRIPOD statement, our adherence assessment can serve as a baseline measurement. We recommend in all subsequent evaluations of adherence to TRIPOD to use our TRIPOD adherence assessment form, including detailed scoring rules. This form is freely accessible through the TRIPOD website and the EQUATOR website, and can be used by anyone (e.g., researchers, reviewers, editors) evaluating adherence to TRIPOD.

A question that arises is who is responsible for accurate reporting of biomedical research. In the first place, based on ethical principles for medical research, researchers are. The Declaration of Helsinki states that "researchers have a duty to make publicly available the results of their research on human subjects and are accountable for the completeness and accuracy of their reports".[4] However, it is also stated that (besides researchers) authors, sponsors, editors and publishers all have ethical obligations, that all parties should follow the accepted guidelines for ethical reporting, and that research reports not in accordance with these principles should not be accepted for publication. Hence, research reporting is a shared responsibility of all stakeholders involved.

Notwithstanding the clear ethical obligations, apparently there are barriers with regard to adequate reporting and the use of reporting guidelines, since, despite small improvements, reporting is still considered suboptimal in many fields.[5] A recent scoping review identified a variety of strategies targeting various stakeholders, that were developed to address potential barriers and improve adherence to reporting

guidelines, including strategies applied during the writing phase (structured formats and automated tools), at manuscript submission (endorsement and other editorial actions, e.g. offering authors personal assistance with manuscript preparation), or during peer review (compliance checking).[6]

Reporting guidelines can only have an impact on the completeness of reporting if potential users are aware of their existence. Endorsement of reporting guidelines by medical journals in the form of mentioning these guidelines in their instructions to authors is a strategy to enhance awareness. Our evaluation of online instructions to authors of 337 medical journals showed that almost two thirds endorse one or more reporting guidelines. There is, however, room for improvement with regard to the formulation of recommendations. Despite the evidence that requiring adherence to reporting guidelines does improve completeness of reporting,[7] instructions to authors are often not that directive but more suggestive. About half of the editors that participated in our survey stated that authors must submit a checklist alongside their manuscript or provide a statement that they followed a reporting guideline, however, only few survey participants mentioned that editors or peer reviewers check whether manuscripts indeed comply to the reporting guideline.

Another way to improve author knowledge on reporting and reporting guideline is by providing education and training. Evidence regarding effectiveness of these types of interventions on improving the practical use and understanding of reporting guidelines is, however, scarce.[6]



Further development, evaluation, and implementation of strategies to improve research reporting are needed and these should target all stakeholders mentioned before. Lack of knowledge and awareness among various stakeholders is thought to be an important barrier to using reporting guidelines, also by respondents in our editor survey (Chapter 5).[8-10] Education and training, for example, should not only be developed for (early career) researchers, but also for peer reviewers, editorial staff, publishers, and funders. Again, considering the shared responsibility for adequate research reporting, it is not up to a single stakeholder to take initiative. The potential danger of this shared responsibility is that everyone is looking at each other and no one takes action. Therefore, an international coordinating initiative is indispensable. The EQUATOR Network has that coordinating role and provides various tools for authors, peer reviewers, and editors. Its website contains an extensive database of over 400 existing reporting guidelines including a tool to select the appropriate reporting guideline. Still, several editors participating in our survey indicated that the

website needs a revision to make it more user friendly. Apart from the stakeholders mentioned before, EQUATOR also provides guidance for developers of reporting guidelines. This guideline for reporting guideline developers lists 18 recommended steps for developing a health research reporting guideline.[11] It does not present a standardized format. In addition, the EQUATOR database currently is inclusive and does not exclude guidelines based on their development methods.[12] Addressing this could be a potential future way to restrict the enormous number of available guidelines and facilitate more uniformity in formats and terminology among the various reporting guidelines.

Next steps to facilitate the uptake of TRIPOD and thereby enhance completeness of reporting of prediction model studies, should include the provision of training and the development of (online) educational tools, accessible for all relevant stakeholders, i.e. students, researchers, peer reviewers, journal editors, clinicians, and funders. Ideally, the development of training materials and educational tools is informed by input from the relevant stakeholder. For this purpose, a survey among authors and peer reviewers would provide useful information and is one of the future initiatives to undertake.

Furthermore, a translation of our TRIPOD adherence assessment form (Chapter 3) into an automated tool would be useful not only to researchers performing adherence assessments, but also to medical journals or peer reviewers when checking compliance of manuscripts to TRIPOD. Finally, to follow the uptake of TRIPOD, a follow-up adherence assessment is planned.

For dissemination and implementation of tools and findings we will collaborate with the EQUATOR Network and Cochrane. Involving these two international organizations will help to increase the impact on adherence to TRIPOD and to achieve our aim of transparent and accurate research reports concerning prediction models, making them more usable in clinical practice.

*De-implementation of low-value care*

Low value healthcare practices may cause harm to patients and can lead to inefficient use of limited healthcare resources. In addition, estimates of the prevalence of low-value care range from 10% to 30%, however, for specific healthcare practices estimates up to 89% have been reported.[13-17] As a result, there is a growing interest in low-value healthcare practices and strategies to reduce them. While the field of implementation science has produced a wealth of theories and evidence on promoting the implementation of (new) healthcare practices in general, the specific challenge

of reducing low-value practices (de-implementation) is only starting to receive attention.[18] There are parallels between implementation and de-implementation, as both require patients, healthcare providers, or other actors to change their behaviour, however, also concepts unique to de-implementation are presumed, which need to be systematically explored [18-20] Recently, several frameworks were introduced addressing the concept of de-implementation.[18,21-25]

There are challenges regarding the design of a de-implementation strategy. The starting point of a de-implementation initiative is the recognition that there is a healthcare practice of low-value ('low-value care') and a desire to remove, replace, reduce or restrict it. The certainty of the evidence regarding this low-value care is a potential factor that influences the de-implementation process, however, to what extent needs to be determined.[21,24,25] Furthermore, low-value care can vary in complexity from a single test or medication to a complex of several interventions involving specific skills or resources. In addition, there are characteristics on the level of healthcare providers, patients, and organisations that can act as barriers or facilitators to de-implementation. Therefore, when designing a strategy to change behaviour all these potential influencing factors should be taken into account and explored in advance. Very few de-implementation studies, however, reported to have explored such factors before determining the strategy targets (provider, patient, organisation) and interventions. Our finding that attitude rather than knowledge seems to be a driver for providing low value care suggests that it is worthwhile to consider other interventions than just provider education. Our review of de-implementation strategies also indicates the potential effectiveness of including other interventions and targeting patients. Several of the de-implementation frameworks underline the importance of the patient-healthcare provider interaction in the process of de-implementation.[22,23,25]

Also the evaluation of the impact of a de-implementation strategy is challenging. First this requires measurement of low-value care, which is not easy, as it needs clear definitions of what is considered appropriate and inappropriate care.[16,26-28] In addition, available data should be detailed enough to determine (in)appropriateness, which might not be the case for routinely collected data. In our set of de-implementation studies, 37% was able to present results for actual inappropriate care, others used total volume (combination of appropriate and inappropriate) instead. Also, a national program in the Netherlands called "To do or not to do?", that evaluated eight diverse de-implementation projects, reported difficulties in data collection as a barrier. [29] A second challenge in the evaluation of de-implementation strategies is the choice of

outcomes. Although the ultimate aim of de-implementation is a permanent change in behaviour, only a minority of the included studies also measured long-term effects of their strategy. As de-implementation strategies can have unintended consequences for patients, health professionals, and organizations, it is important to evaluate these.[25,30] The evaluation of outcomes other than utilization is, however, not yet very common, as was concluded in a recent review and also reflects our own experience.[30]

This thesis provides a broad overview of the de-implementation literature. As de-implementation is strongly associated with the context in which it takes place, a logical next step is to explore existing de-implementation studies that focus on a specific low-value healthcare practice. We are currently working on a systematic review of strategies to reduce inappropriate proton pump inhibitor use for stress ulcer prophylaxis in hospitalized, non-intensive care unit patients, of which the abstract with preliminary findings is presented in the Box to indicate the challenges we have encountered.

**Box. Case study (work in progress): Reducing inappropriate use of Proton Pump Inhibitors (PPI) for Stress Ulcer Prophylaxis (SUP) in hospitalized patients, a systematic review**

***Objective***
To identify and compare strategies to reduce the use of inappropriate proton pump inhibitors (PPI) for stress ulcer prophylaxis (SUP) in hospitalized, non-intensive care unit (non-ICU) patients.

***Methods***
MEDLINE and Embase databases were searched on the 8th of January, 2020.

Eligible studies included adult, hospitalized patients in non-ICU settings who were receiving PPI for SUP, and evaluated an intervention to reduce the use of inappropriate PPI. Randomized trials and comparative observational studies (including interrupted time-series (ITS) and controlled before-after studies with or without a parallel control group) were eligible.

Included studies were critically appraised using criteria developed by the Cochrane Effective Practice and Organisation of Care (EPOC) group, that were adapted to fit the eligible study designs.

Besides the primary outcome (inappropriate PPI prescription or use), additional outcomes of interest included pharmaceutical effects (symptoms of acid reflux; ulcer and upper gastrointestinal bleeding), adverse pharmaceutical effects (diarrhoea or obstipation, abdominal pain, *Clostridium difficile* infections, hospital-acquired pneumonia, electrolyte disturbances), and healthcare use (e.g. length of stay (LOS), ICU or hospital admission, emergency department (ED) visit, alternative medication use).

The review protocol was registered in PROSPERO (January 14th, 2020; acknowledgement of receipt number 165508)

***Results***

After screening 1863 references, ten studies met the inclusion criteria.

*Characteristics of included studies*

Apart from one non-randomized trial, all studies had a before-after design without a parallel control group. Studies were conducted at one (n=4) or multiple (n=6) wards in a general (n=5) or academic hospital (n=3), or both (n=2).

The definition of inappropriate PPI use was based on literature or (inter)national guidelines in all but one study, in which a panel of experts defined inappropriateness. The literature and guidelines used varied between studies and only partially overlapped.

One study identified barriers and facilitators to reducing PPI use prior to designing the de-implementation strategy. Three other studies referred to literature for effective de-implementation or teaching strategies. Three de-implementation strategies addressed the medical staff, six the medical staff and the organisation, and one the pharmacy staff and organisation. All de-implementation strategies contained an educational component (meetings and/or materials), in combination with an organizational intervention (n=7), reminders (n=2), and audit feedback (n=3), except for one strategy that combined reminders with an organizational intervention.

Eight studies evaluated the effect of the strategy on inappropriate care and two studies on overall volume (combined appropriate and inappropriate care). Secondary outcomes that were measured were pharmaceutical effects (n=1), adverse pharmaceutical effects (n=2), and LOS (n=3).

*Critical appraisal*

The non-randomized, observational study designs without parallel control group lead to a risk of bias in the included studies. Differences between baseline characteristics of analysed groups were identified in two studies. In addition, information to judge blinding of outcome assessment and selective reporting was lacking for most studies.

*Effectiveness of strategies*
Six studies evaluated the effect of the strategy on new PPI prescriptions. Baseline proportion of inappropriate PPI prescriptions ranged from 19% to 92% and three studies found a significant reduction after de-implementation.
Five studies evaluated the effect on PPI use (also taking inappropriate duration or dosage into account). Baseline proportion of inappropriate PPI prescriptions ranged from 43% to 82% and four studies found a significant reduction after de-implementation. One study presented a reduction in PPI prescriptions over a period of three years. None of the other studies provided long term results.

With regard to the secondary outcomes, no significant differences in the occurrence of pharmaceutical effects (n=1 study) and in LOS (n=3) were seen. One study found a significant decrease of adverse pharmaceutical effects after de-implementation, whereas the other study found no significant difference.

No pooled effect estimate was calculated because of heterogeneity.

This case study illustrates that despite the relatively narrow focus of the review on one specific low-value healthcare practice, the resulting set of included studies is still quite heterogeneous. Differences between studies regarding population and setting and defining low-value PPI likely explain the variation in baseline proportions of PPI prescriptions and use. Furthermore, studies used different combinations of interventions and outcomes.

This empirical example underlines the need for process evaluations to assess whether a strategy was implemented as intended and which barriers were encountered.[31] A process evaluation can help to explain the observed effects of a de-implementation strategy in the light of all complexities arising from the multiple interacting components and factors. Process evaluations can also provide insight in how de-implementation differs from implementation. Our review (Chapter 7) identified two studies which performed a process evaluation.[32,33] In addition, the Dutch "To do or not to do?" program used process evaluations in evaluating eight de-implementation projects, which revealed useful information to roll out successful strategies at a larger scale .[29]

Authors of de-implementation studies should start with taking all the contextual factors into account when designing a strategy. Subsequently, in the evaluation of the impact of the de-implementation strategy they need to collect and report all essential information needed to interpret and apply their results into practice. This includes knowledge on barriers and facilitators, de-implementation strategy details, sustainability of observed effects, and insight into unintended consequences of the

de-implementation strategy. There are several relevant reporting guidelines that can assist authors therein.[34-36]

**Concluding remarks**

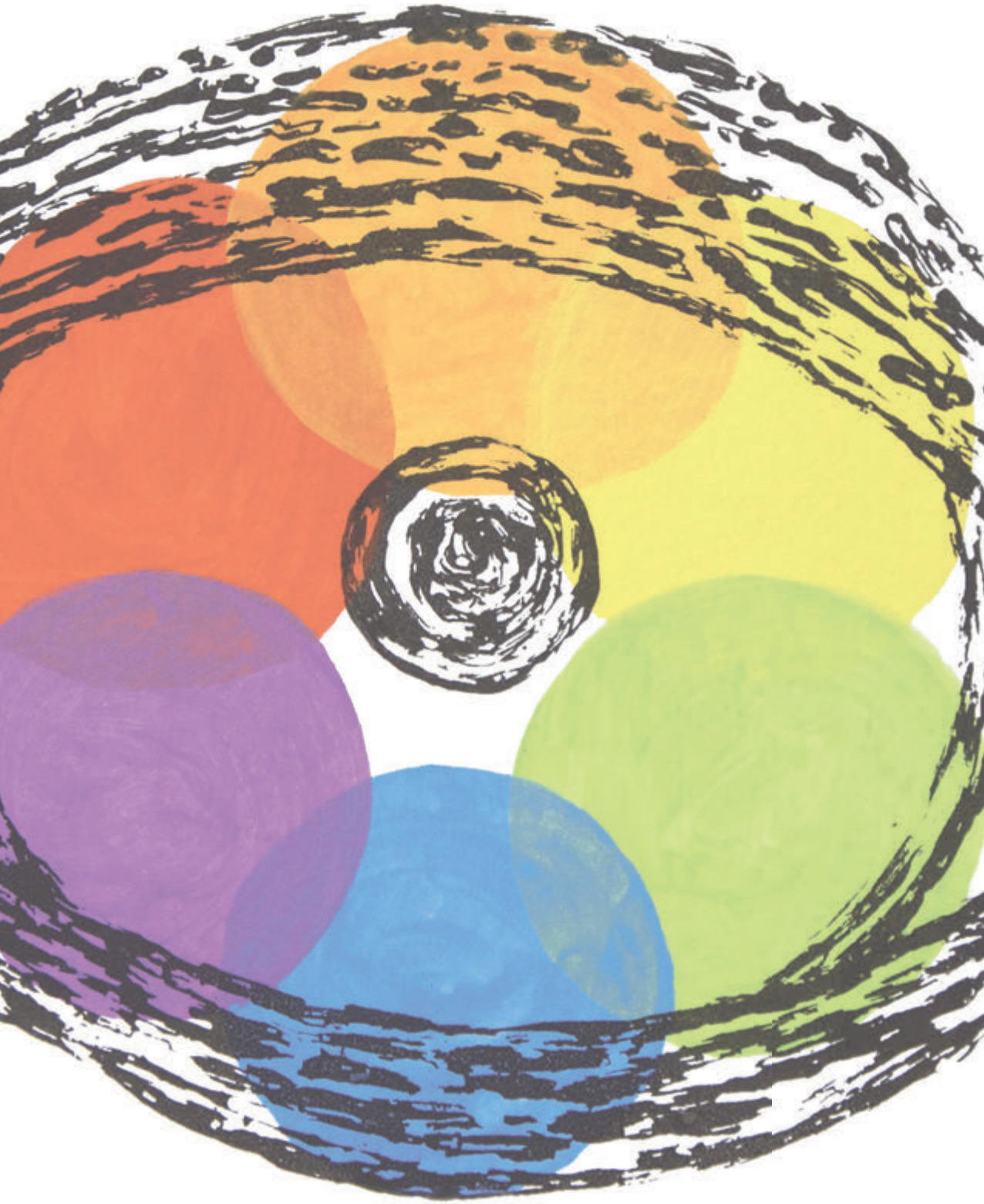
Facilitating EBM by maximizing the value and promoting the use of available evidence in clinical decision making is the main theme of this thesis. If research findings do not find their way to routine clinical practice, they cannot benefit patients. We explored two concepts contributing to maximizing the value of research and moving findings into practice: transparent and accurate reporting and implementation. This thesis addressed these two concepts in separate sections. However, there are clear overlapping aspects: to promote the use of reporting guidelines, one needs the principles of implementation and a better understanding of the process of (de-) implementation cannot be realized without complete and transparent reporting of the research methods and findings. A mixed-methods approach with a combination of quantitative and qualitative research will provide the opportunity to collect all the necessary information to design and evaluate effective strategies to promote the uptake of research findings in clinical practice and let patients benefit from the available evidence.

## References


1. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev.* 2012;11:Mr000030.
2. Korevaar DA, Wang J, van Enst WA, Leeflang MM, Hooft L, et al. Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. *Radiology.* 2015;274(3):781-789.
3. Sekula P, Mallett S, Altman DG, Sauerbrei W. Did the reporting of prognostic studies of tumour markers improve since the introduction of REMARK guideline? A comparison of reporting in published articles. *PLoS One.* 2017;12(6):e0178531.
4. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191-2194.
5. Jin Y, Sanger N, Shams I, Luo C, Shahid H, et al. Does the medical literature remain inadequately described despite having reporting guidelines for 21 years? - A systematic review of reviews: an update. *J Multidiscip Healthc.* 2018;11:495-510.
6. Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. Scoping review on interventions to improve adherence to reporting guidelines in health research. *BMJ Open.* 2019;9(5):e026589.
7. Agha RA, Fowler AJ, Limb C, Whitehurst K, Coe R, et al. Impact of the mandatory implementation of reporting guidelines on reporting quality in a surgical journal: A before and after study. *Int J Surg.* 2016;30:169-172.
8. Fuller T, Pearson M, Peters J, Anderson R. What affects authors' and editors' use of reporting guidelines? Findings from an online survey and qualitative interviews. *PLoS One.* 2015;10(4):e0121585.
9. Grindlay DJ, Dean RS, Christopher MM, Brennan ML. A survey of the awareness, knowledge, policies and views of veterinary journal Editors-in-Chief on reporting guidelines for publication of research. *BMC Vet Res.* 2014;10:10.
10. Shamseer L, Hopewell S, Altman DG, Moher D, Schulz KF. Update on the endorsement of CONSORT by high impact factor journals: a survey of journal "Instructions to Authors" in 2014. *Trials.* 2016;17(1):301.
11. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med.* 2010;7(2):e1000217.
12. Altman DG, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. *J R Soc Med.* 2016;109(2):67-77.
13. Brownlee S, Chalkidou K, Doust J, Elshaug AG, Glasziou P, et al. Evidence for overuse of medical services around the world. *Lancet.* 2017.

14. Colla CH, Mainor AJ, Hargreaves C, Sequist T, Morden N. Interventions Aimed at Reducing Use of Low-Value Health Services: A Systematic Review. *Med Care Res Rev.* 2017;74(5).
15. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet.* 2003;362(9391):1225-1230.
16. Korenstein D, Falk R, Howell EA, Bishop T, Keyhani S. Overuse of health care services in the United States: an understudied problem. *Arch Intern Med.* 2012;172(2):171-178.
17. Morgan DJ, Brownlee S, Leppin AL, Kressin N, Dhruva SS, et al. Setting a research agenda for medical overuse. *BMJ.* 2015;351:h4534.
18. Grimshaw JM, Patey AM, Kirkham KR, Hall A, Dowling SK, et al. De-implementing wisely: developing the evidence base to reduce low-value care. *BMJ Qual Saf.* 2020.
19. Scott IA, Elshaug AG. Foregoing low-value care: how much evidence is needed to change beliefs? *Intern Med J.* 2013;43(2):107-109.
20. van Bodegom-Vos L, Davidoff F, Marang-van de Mheen PJ. Implementation and de-implementation: two sides of the same coin? *BMJ Qual Saf.* 2016.
21. Prasad V, Ioannidis JP. Evidence-based de-implementation for contradicted, unproven, and aspiring healthcare practices. *Implement Sci.* 2014;9:1.
22. Morgan DJ, Leppin AL, Smith CD, Korenstein D. A Practical Framework for Understanding and Reducing Medical Overuse: Conceptualizing Overuse Through the Patient-Clinician Interaction. *J Hosp Med.* 2017;12(5):346-351.
23. Parchman ML, Henrikson NB, Blasi PR, Buist DS, Penfold R, et al. Taking action on overuse: Creating the culture for change. *Healthc (Amst).* 2017;5(4):199-203.
24. Norton WE, Chambers DA, Kramer BS. Conceptualizing De-Implementation in Cancer Care Delivery. *J Clin Oncol.* 2019;37(2):93-96.
25. Norton WE, Chambers DA. Unpacking the complexities of de-implementing inappropriate health interventions. *Implement Sci.* 2020;15(1):2.
26. Baker DW, Qaseem A, Reynolds PP, Gardner LA, Schneider EC. Design and use of performance measures to decrease low-value services and achieve cost-conscious care. *Ann Intern Med.* 2013;158(1):55-59.
27. Chan KS, Chang E, Nassery N, Chang HY, Segal JB. The state of overuse measurement: a critical review. *Med Care Res Rev.* 2013;70(5):473-496.
28. de Vries EF, Struijs JN, Heijink R, Hendrikx RJ, Baan CA. Are low-value care measures up to the task? A systematic review of the literature. *BMC Health Serv Res.* 2016;16(1):405.
29. Dulmen Sv, Heus P, Kool T, Verkerk E. *Doen of laten in de gezondheidszorg. Een onderzoek naar de mogelijkheden van terugdringen van niet-gepaste zorg.* Nijmegen: IQ healthcare;2019.

30. Maratt JK, Kerr EA, Klamerus ML, Lohman SE, Froehlich W, et al. Measures Used to Assess the Impact of Interventions to Reduce Low-Value Care: a Systematic Review. *J Gen Intern Med.* 2019;34(9):1857-1864.
31. Moore GF, Audrey S, Barker M, Bond L, Bonell C, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ.* 2015;350:h1258.
32. Trietsch J, van Steenkiste B, Hobma S, Frericks A, Grol R, et al. The challenge of transferring an implementation strategy from academia to the field: a process evaluation of local quality improvement collaboratives in Dutch primary care using the normalization process theory. *J Eval Clin Pract.* 2014;20(6):1162-1171.
33. Voorn VMA, Marang-van de Mheen PJ, van der Hout A, Hofstede SN, So-Osman C, et al. The effectiveness of a de-implementation strategy to reduce low-value blood management techniques in primary hip and knee arthroplasty: a pragmatic cluster-randomized controlled trial. *Implement Sci.* 2017;12(1):72.
34. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, et al. Standards for Reporting Implementation Studies (StaRI) Statement. *BMJ.* 2017;356:i6795.
35. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf.* 2016;25(12):986-992.
36. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ.* 2014;348:g1687.



# Appendices



## Summary

The term evidence-based medicine (EBM) refers to the process of integrating the available information from clinical research (evidence) with clinical expertise and patient preferences in making decisions about the care of individual patients. EBM requires researchers to write useful reports of their research, in which the research question, methods, and results and their implications are clearly described. Adequate reporting, however, is not enough. Usually, additional activities are needed to ensure the uptake of research evidence in routine clinical practice. The aim of this thesis is to explore and improve the methods to report healthcare research and implement research findings, which are both essential components to facilitate EBM.

The first part of this thesis focuses on the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement, a guideline that aims to improve the adequacy of reporting of prediction model studies. In **Chapter 2** we describe the assessment of the completeness of reporting of prediction model studies published just before the introduction of the TRIPOD statement. We searched for prediction model studies in the top ten impact factor journals within each of 37 clinical domains. Diagnostic or prognostic prediction model studies of all types (development, external validation, and added value of new predictors to existing models) were eligible. We evaluated 170 prediction models and concluded that, in general, they were poorly reported, as more than half of the items that are considered essential according to TRIPOD were not or not fully addressed. Information essential for identification of prediction model studies, use of a model in individual risk prediction, or for external validation of a prediction model was often not detailed enough. Aspects of prediction model studies that require improved reporting are title, abstract, statistical analysis methods, and results (i.e. model specifications and model performance). Our findings enable targeted training, education and guidance for authors, researchers, and journal editors.

For the assessment presented in Chapter 2 we transformed the original 22 items of the TRIPOD statement into a systematic and transparent adherence assessment form. In **Chapter 3**, aiming to promote uniformity in measuring adherence to TRIPOD, we share our experiences with designing this form and creating TRIPOD adherence scoring rules. Challenges encountered specific to TRIPOD were the existence of different types of prediction model studies and possible combinations of these within publications. More general issues included dealing with multiple reporting elements, reference to information in another publication, and nonapplicability of items. We recommend our adherence assessment form to be used by anyone (e.g., researchers, reviewers,

and editors) evaluating adherence to TRIPOD, to make assessments consistent and comparable over time and between clinical domains. In general, when developing a form to assess adherence to a reporting guideline, we recommend formulating specific adherence elements (if needed multiple per reporting guideline item) using unambiguous wording and the consideration of issues of applicability in advance.

Following our findings of incomplete reporting of titles and abstracts (Chapter 2), in **Chapter 4** we present a checklist and corresponding guidance for reporting prediction model studies in abstracts. This checklist was developed using a modified Delphi procedure in the form of a web-based survey among 110 experts in the field of prediction modeling. Based on items of the TRIPOD statement and existing reporting guidelines for abstracts, a list of 32 potentially relevant items was the starting point of this survey. After three survey rounds there was consensus on the items that should be considered as the minimum set of information that is required for informative abstracts on prediction models. TRIPOD for Abstracts is a checklist of 12 items that is applicable to all types of prediction model studies (including development, external validation, added value and model updating studies), regardless the clinical domain or the statistical approach used. In combination with the explanation and examples of adequate reporting we provided, it will contribute to improved reporting, and thereby facilitate readers and reviewers in identifying a potentially relevant prediction model study, as well as judgeing its relevance and validity.

**Chapter 5** focuses on the endorsement of TRIPOD and other reporting guidelines by medical journals and on journal editors' opinions and experiences regarding promoting the use of reporting guidelines. We searched the online 'Instructions to authors' of 337 journals of various clinical domains and invited the editors-in-chief to participate in an online survey. We found that almost two thirds of medical journals endorse one or more reporting guidelines and TRIPOD was endorsed by 9%. None of the TRIPOD endorsing journals had made the use of TRIPOD mandatory. Lack of knowledge among authors, reviewers, and editors; putting a burden on authors and peer reviewers; inflexibility; fear of less submissions; and the large number of available reporting guidelines, were identified as potential barriers to using them. Editors of medical journals suggested the following to overcome these barriers: make adherence to or use of reporting guidelines mandatory for authors and reviewers; education and dissemination of tools how to use the reporting guideline; and the use of software applications and automated tools for identifying reporting guidelines and checking publications on guideline adherence. This chapter provides insight in the journal's editorial policies regarding reporting guidelines, and on (potential) barriers

and facilitators to endorsement and active use of these guidelines. This information can be used to develop targeted initiatives to promote the use of TRIPOD and other reporting guidelines.

The second part of this thesis addresses the implementation of evidence recommending to abandon the routine use of a specific healthcare practice of low-value, so called de-implementation. **Chapter 6** describes a synthesis of the existing evidence regarding potential barriers and facilitators to de-implementation in healthcare settings. We systematically searched for relevant studies and performed a qualitative evidence synthesis using an existing framework for grouping barriers and facilitators for change. We identified 404 unique factors (either barrier or facilitator) across the 111 included articles. For a large part these factors were related to the individual healthcare provider, and rather to attitude, than to knowledge or behaviour. Besides healthcare provider factors, factors related to the patient, social context, organizational context and economical/political context were identified. Although future research should investigate this more specifically, it seems that patient-provider interaction, the fear of consequences of withholding a test or treatment, and financial incentives are more important factors in de-implementation than in implementation. This qualitative evidence synthesis provides insight into the range of factors affecting the success of strategies to reduce low-value care, which knowledge can advance the design of these strategies.

In **Chapter 7** we show the results of a systematic review to compare the effectiveness of various de-implementation strategies and to identify characteristics associated with their success. We included 49 randomized controlled trials evaluating a de-implementation strategy and found that, compared to usual care, de-implementation led to a median relative reduction in the use of a low-value healthcare practice of 13%. The effect tended to be smaller for strategies consisting of a single intervention. To reduce therapeutic low-value care services, a strategy targeted at patients was inclined to achieve a larger effect compared to strategies that did not address patients. Strategies containing audit and feedback showed a trend towards a larger effect than strategies without this intervention and incorporating reminders seemed beneficial for strategies addressing diagnostic healthcare practices. Details regarding perceived barriers and facilitators, sustainability of effect and potential (unintended) consequences of reducing a low-value care practice on patient health or healthcare use were often not provided. This contextual information is, however, essential for interpretation and application of the results and for rolling out successful strategies at a larger scale.

As Chapter 7 provides a broad overview of the de-implementation literature and the included studies used many different combinations of interventions to reduce low-value care with little overlap, we concentrate in **Chapter 8** on a more specific type of healthcare practice and setting. In this systematic review we included randomized controlled trials evaluating strategies to reduce the use of medical tests in primary care. Despite the more narrow focus, there was still substantial heterogeneity among the included studies in terms of type and role of medical tests, components and targets of intervention. As a result, it was difficult to disentangle the effect of each of these factors. Results and conclusions were analogous to those in Chapter 7: 11 of 16 included strategies (69%) showed an effect and the median relative reduction in the use of low-value medical tests was 17%. Especially strategies consisting of multiple components, including reminders, audit and feedback, or patient-targeted interventions, showed a larger effect than those without these components. Yet, to widely implement these strategies in primary care settings, future studies need to investigate sustainability of the effect, adverse events, cost-effectiveness, and patient-reported outcomes.

This thesis concludes with a general discussion (**Chapter 9**) elaborating on the lessons learned and the implications for practice and research. Challenges discussed with regard to reporting and the use of reporting guidelines, are the shared responsibility of the various stakeholders involved, and the importance of developing, evaluating, and implementing strategies to improve research reporting. Possible next steps to facilitate the uptake of TRIPOD and thereby enhance completeness of reporting of prediction model studies, are the provision of training, design of (online) educational tools, and the development of an automated tool based on our TRIPOD adherence assessment form (Chapter 3), preferably in collaboration with the TRIPOD Steering Group, the EQUATOR Network and Cochrane. Challenges regarding de-implementation that were discussed in Chapter 9 are mainly related to the context in which de-implementation takes place. This context should be taken into account in the design, as well as in the evaluation and reporting of strategies aimed at reducing the routine use of a specific healthcare practice. This is illustrated in the empirical example presented.

Both concepts explored in this thesis show that a combination of quantitative and qualitative research is needed to collect all the necessary information to design and evaluate effective strategies to promote the uptake of research findings in clinical practice. Only then we can let patients benefit from the available evidence and maximize the value of our research.

## Samenvatting

De term *evidence-based medicine* (EBM) verwijst naar het proces waarbij beschikbare informatie uit klinisch-wetenschappelijk onderzoek (*evidence*) wordt geïntegreerd met klinische expertise en patiëntvoorkeuren bij het nemen van beslissingen voor individuele patiënten. EBM vereist dat onderzoekers op adequate wijze verslag uitbrengen van hun onderzoek, waarbij ze de onderzoeksvraag, methoden, resultaten en implicaties helder beschrijven. Goede verslaglegging is echter niet voldoende. Meestal zijn aanvullende activiteiten nodig om ervoor te zorgen dat onderzoeksresultaten ook daadwerkelijk in de dagelijkse klinische praktijk worden gebruikt. Het doel van dit proefschrift is het verkennen en verbeteren van methoden voor 1) de rapportage van gezondheidsonderzoek en 2) de implementatie van onderzoeksresultaten, welke beide essentiële elementen zijn om EBM te kunnen toepassen.

Het eerste deel van dit proefschrift richt zich op *TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis)*, een richtlijn die tot doel heeft de volledigheid van publicaties over predictiemodellen (voorspellingsmodellen) te verbeteren. In **Hoofdstuk 2** beschrijven we de volledigheid van de verslaglegging van dit type onderzoek voorafgaand aan het verschijnen van TRIPOD. We zochten naar publicaties over predictiemodelonderzoeken in de 10 meest toonaangevende medische tijdschriften binnen elk van 37 medische vakgebieden. Alle typen diagnostisch of prognostisch predictiemodelonderzoek (modelontwikkeling, externe validatie en bepaling van de meerwaarde van het opnemen van nieuwe voorspellers in bestaande modellen) kwamen in aanmerking. We beoordeelden 170 predictiemodellen en kwamen tot de conclusie dat deze over het algemeen slecht werden beschreven, aangezien meer dan de helft van de items die volgens TRIPOD essentieel zijn, niet of niet volledig werden vermeld. Informatie die nodig is voor het traceren van potentieel relevant predictiemodelonderzoek, voor het gebruik van een model voor risicovoorspelling voor een individuele patiënt of voor externe validatie van een predictiemodel, was vaak niet gedetailleerd genoeg beschreven. Aspecten van predictiemodelonderzoek die beter beschreven zouden moeten worden, zijn de titel, het abstract, de statistische analysemethoden en resultaten (d.w.z. modelspecificaties en -prestaties). Onze bevindingen maken gerichte training, opleiding en begeleiding van auteurs, onderzoekers en tijdschriftredacteuren mogelijk.

Voor de beoordeling van de predictiemodellen in Hoofdstuk 2 hebben we voor de 22 oorspronkelijke TRIPOD-items regels opgesteld, aan de hand waarvan we op systematische en transparante wijze de naleving van TRIPOD konden beoordelen.

In **Hoofdstuk 3** beschrijven we hoe we dat gedaan hebben met als doel uniformiteit te bevorderen in het meten van de volledigheid van de verslaglegging volgens TRIPOD. TRIPOD-specifieke uitdagingen hierbij waren de verschillende typen predictiemodelonderzoeken die er zijn, en de mogelijke combinaties waarin deze kunnen voorkomen binnen publicaties. Meer algemene kwesties waren hoe om te gaan met items die uit meerdere elementen bestaan, met verwijzingen naar informatie in een andere publicatie en met het niet van toepassing zijn van items. We bevelen aan dat iedereen die de volledigheid van de verslaglegging van predictiemodelonderzoek wil evalueren (bijv. onderzoekers, reviewers en redacteuren), ons beoordelingsformulier gebruikt om dergelijke evaluaties zo over de tijd en tussen verschillende medische vakgebieden consistent en vergelijkbaar te maken. In het algemeen adviseren we om bij het ontwikkelen van een formulier om de naleving van een rapportagerichtlijn te beoordelen, specifieke beoordelingscriteria te formuleren (indien nodig meerdere per item van de rapportagerichtlijn) en daarbij ondubbelzinnige bewoordingen te gebruiken en vooraf na te denken over hoe om te gaan met items die mogelijk niet in alle gevallen van toepassing zijn.

In Hoofdstuk 2 constateerden wij dat de rapportage van titel en abstract vaak onvolledig is. Naar aanleiding daarvan introduceerden we in **Hoofdstuk 4** een checklist met bijbehorende uitleg voor het rapporteren van abstracts van predictiemodelonderzoeken. Deze checklist werd ontwikkeld met behulp van een Delphi-procedure in de vorm van een online enquête onder 110 experts op het gebied van predictiemodellen. Op basis van items van TRIPOD en bestaande rapportagerichtlijnen voor abstracts werd een lijst van 32 potentieel relevante items opgesteld, die diende als uitgangspunt van de Delphi-procedure. Na drie enquêterondes was er consensus over een minimale set van benodigde gegevens voor een informatief abstract van predictiemodelonderzoek. *TRIPOD for Abstracts* is een checklist van 12 items die van toepassing is op alle typen predictiemodelonderzoek (inclusief modelontwikkeling, externe validatie, bepalen van de toegevoegde waarde van voorspellers en modelupdates), ongeacht het medische vakgebied of de gebruikte statistische methode. In combinatie met onze toelichting en voorbeelden van adequate verslaglegging beoogt de checklist bij te dragen aan een betere verslaglegging en daarmee lezers en reviewers te helpen bij het traceren van potentieel relevant predictiemodelonderzoek, evenals bij het beoordelen van de relevantie en validiteit ervan.

In **Hoofdstuk 5** onderzochten we in welke mate TRIPOD en andere rapportagerichtlijnen door medische tijdschriften onderschreven worden en wat de meningen en ervaringen

zijn van redacteuren ten aanzien van het bevorderen van het gebruik van dergelijke richtlijnen. We raadpleegden de online 'Instructions for authors' van 337 tijdschriften uit verschillende medische vakgebieden en nodigden daarnaast de hoofdredacteuren van die tijdschriften uit om deel te nemen aan een online enquête. Bijna tweederde van de medische tijdschriften bleek één of meer rapportagerichtlijnen te onderschrijven. TRIPOD werd door 9% onderschreven en geen van deze tijdschriften had het gebruik ervan verplicht gesteld. Redacteuren van medische tijdschriften noemden de volgende mogelijke belemmeringen voor het gebruik van een rapportagerichtlijn: gebrek aan kennis van rapportagerichtlijnen bij auteurs, reviewers en redacteuren; toename van de werklast voor auteurs en peer-reviewers; beperkte flexibiliteit van rapportagerichtlijnen; vrees voor minder ingediende artikelen; en het bestaan van een te groot aantal van deze richtlijnen. Ze stelden het volgende voor om deze belemmeringen aan te pakken: het gebruik van rapportagerichtlijnen en de naleving ervan verplicht stellen voor auteurs en reviewers, onderwijs geven over het gebruik van een rapportagerichtlijn en bijbehorende hulpmiddelen verspreiden, en het gebruik van softwaretoepassingen voor het identificeren van rapportagerichtlijnen en het controleren van publicaties op naleving van richtlijnen. Dit hoofdstuk geeft inzicht in het redactionele beleid van medische tijdschriften met betrekking tot rapportagerichtlijnen en in (potentiële) belemmerende en bevorderende factoren voor onderschrijving en actief gebruik van deze richtlijnen. Deze informatie kan worden gebruikt om gerichte initiatieven ter bevordering van het gebruik van TRIPOD en andere rapportagerichtlijnen te ontwikkelen.

Het tweede deel van dit proefschrift is gericht op de implementatie van onderzoeksresultaten die aanbevelen bepaald routinematig zorggebruik te staken, zogenoemde deïmplementatie. In **Hoofdstuk 6** presenteren we een overzicht van potentiële belemmerende en bevorderende factoren voor deïmplementatie. We zochten op systematische wijze naar relevante artikelen en vatten deze op kwalitatieve wijze samen aan de hand van een bestaand raamwerk om belemmerende en bevorderende factoren voor verandering in te delen. We vonden 404 unieke factoren (belemmerend of bevorderend) in 111 geselecteerde artikelen. Deze factoren hadden, meer dan op kennis of gedrag, betrekking op de attitude van individuele zorgverleners. Daarnaast vonden we factoren ten aanzien van de patiënt, de sociale context, de organisatorische context en de economische of politieke context. Hoewel toekomstig onderzoek dit specifieker zal moeten nagaan, lijken de interactie tussen patiënt en zorgverlener, angst voor de gevolgen (van het niet uitvoeren van een test of behandeling) en financiële prikkels een grotere rol te spelen bij deïmplementatie van reeds ingeburgerde zorg dan bij de implementatie van nieuwe zorg. Deze kwalitatieve

evidencesynthese geeft inzicht in de reeks factoren die van invloed zijn op het succes van strategieën om zorg zonder meerwaarde terug te dringen, hetgeen kan bijdragen aan het ontwerpen van toekomstige strategieën.

In **Hoofdstuk 7** vergelijken we aan de hand van een systematisch literatuuronderzoek de effectiviteit van verschillende deïmplementatiestrategieën en stellen we kenmerken vast die verband houden met hun succes. We identificeerden 49 gerandomiseerde gecontroleerde onderzoeken waarin een strategie voor deïmplementatie werd geëvalueerd, en vonden dat deïmplementatie leidde tot een mediane relatieve afname van 13% in het gebruik van zorg met weinig toegevoegde waarde. Het effect van strategieën die uit één enkele interventie bestonden, was in het algemeen kleiner. Voor het terugdringen van therapeutische zorg leek een op patiënten gerichte strategie een groter effect te hebben dan strategieën die niet op patiënten waren gericht. Strategieën die *audit and feedback* (terugkoppeling over het handelen op basis van toetsing) als interventie toepasten, neigden naar een groter effect dan strategieën zonder deze interventie, en het opnemen van *reminders* (herinneringen en beslissingsondersteuningen) als interventie leek gunstig voor strategieën die gericht waren op terugdringen van diagnostische handelingen met weinig toegevoegde waarde voor de patiënt. Gedetailleerde informatie over de ervaren belemmerende en bevorderende factoren, het aanhouden van eenmaal opgetreden effecten en mogelijke (onbedoelde) gevolgen voor de gezondheid van patiënten of het zorggebruik werd vaak niet verstrekt. Deze contextuele informatie is echter essentieel voor de interpretatie en toepassing van de resultaten en voor het uitrollen van succesvolle strategieën op grotere schaal.

Aangezien Hoofdstuk 7 een breed overzicht geeft van de deïmplementatieliteratuur en de opgenomen onderzoeken veel verschillende combinaties van interventies gebruikten met weinig overlap, richten we ons in **Hoofdstuk 8** op een bepaald onderwerp in een specifieke setting. Voor dit systematische literatuuronderzoek kwamen gerandomiseerde gecontroleerde onderzoeken in aanmerking die strategieën evalueerden om het gebruik van diagnostische tests van weinig toegevoegde waarde in de eerstelijnszorg te verminderen. Ondanks het nauwere perspectief, was er nog steeds sprake van een grote variatie tussen de bestudeerde onderzoeken qua type en de rol van de terug te dringen diagnostische tests en qua componenten en doelen van de ingezette strategieën. Als gevolg hiervan was het moeilijk om het effect van afzonderlijke factoren te ontrafelen. De resultaten en conclusies van ons systematische literatuuronderzoek kwamen overeen met die in Hoofdstuk 7: 11 van de 16 opgenomen strategieën (69%) toonden een effect en de mediane relatieve afname in het gebruik

van de onnodig geachte diagnostische tests was 17%. Vooral strategieën bestaande uit meerdere componenten, met reminders, audit en feedback of patiëntgerichte interventies, lieten een groter effect zien. Om deze strategieën op grote schaal toe te kunnen passen in de eerstelijnszorg, dienen toekomstige onderzoeken het aanhouden van het effect, bijwerkingen, kosteneffectiviteit en door de patiënt gerapporteerde uitkomsten te bestuderen.

Dit proefschrift eindigt met een algemene discussie (**Hoofdstuk 9**) waarin de implicaties van de voorgaande bevindingen voor de praktijk en onderzoek worden uitgewerkt. De uitdagingen die worden besproken met betrekking tot de rapportage en het gebruik van rapportagerichtlijnen, zijn de gedeelde verantwoordelijkheid van de verschillende betrokken belanghebbenden en het belang van het ontwikkelen, evalueren en implementeren van strategieën om rapportage van onderzoek te verbeteren. Mogelijke vervolgstappen om het gebruik van TRIPOD te bevorderen en daardoor de onderzoeksrapportages van predictiemodellen vollediger te maken, zijn het geven van trainingen, het ontwerpen van (online) onderwijsmaterialen en het omzetten van ons beoordelingsformulier om de naleving van TRIPOD te bepalen (Hoofdstuk 3) in een geautomatiseerd instrument. Bij voorkeur doen we dit in samenwerking met de TRIPOD *Steering Group*, het *EQUATOR Network* en Cochrane. De uitdagingen omtrent deïmplementatie houden voornamelijk verband met de context waarin deïmplementatie plaatsvindt. Met deze context moet rekening worden gehouden bij zowel het ontwerp als bij de evaluatie en de rapportage van strategieën om het routinematig gebruik van bepaalde zorg met een beperkte meerwaarde terug te dringen. Dit wordt nog eens benadrukt met een uitgewerkt voorbeeld.

Beide aspecten van EBM die in dit proefschrift worden onderzocht, laten zien dat een combinatie van kwantitatief en kwalitatief onderzoek vereist is om alle informatie te verzamelen die nodig is voor het ontwerpen en evalueren van effectieve strategieën om de opname van onderzoeksresultaten in de klinische praktijk te bevorderen. Alleen dan kunnen we patiënten laten profiteren van de beschikbare evidence en de waarde van ons onderzoek maximaliseren.

## Dankwoord

*"Life is what happens to you while you are busy making other plans"*

Bovenstaande zin uit een door John Lennon geschreven liedje mag dan enigszins cliché zijn, er zit - zoals vaker het geval is met clichés - een kern van waarheid in en de tekst is zeker van toepassing op de totstandkoming van dit proefschrift. Na een weloverwogen keuze voor de opleidingen optometrie en orthoptie was de universiteit namelijk aanvankelijk ver weg. Vele jaren later kwam ik er toch terecht en startte zelfs met een promotietraject. Tijdens mijn promotietraject bleken deadlines soms bij nader inzien minder heilig of haalbaar en liepen dingen anders dan gedacht. En in de allerlaatste fase schopte de COVID-19-pandemie de plannen nog een keer in de war. Het feit dat ik nu een dankwoord op papier zet, betekent dat het proefschrift er ligt en het einde van het promotietraject in zicht is! Graag wil ik iedereen bedanken die daaraan op enigerlei wijze heeft bijgedragen.

Allereerst gaat mijn dank uit naar de mensen die direct betrokken waren bij de totstandkoming van dit proefschrift, te beginnen bij mijn (co)promotoren.

Prof. dr. R.J.P.M. Scholten, beste Rob, wat ben ik blij dat je het aandurfde om een optometrist/orthoptist met de MSc EBP op zak aan je team toe te voegen! Vanaf het begin voelde ik me thuis bij het Dutch Cochrane Centre en van optometrist/orthoptist werd ik steeds meer een klinisch epidemioloog. Je leerde me onderzoek in perspectief te plaatsen met oog voor de klinische toepasbaarheid en dit zorgvuldig over te brengen in een tekst of tijdens het onderwijs. Heerlijk dat je daarbij ook zo kritisch let op correct gebruik van de Nederlandse taal! Niet alleen in de samenwerking, maar ook als persoon waardeer ik je enorm: je betrokkenheid, je humor en de goede gesprekken tijdens onze fietstochtjes naar het station en in de trein. Dat de verdediging van dit proefschrift samenvalt met jouw allerlaatste werkdag voor je pensioen, maakt het extra bijzonder. Zodra het weer kan, nemen we daar alsnog een drankje op, uiteraard mét bitterballen erbij!

Prof. dr. K.G.M. Moons, beste Carl, ik wil je bedanken voor de mogelijkheid die je me bood om een promotietraject te starten. Een project rondom 'jouw' TRIPOD was mijn eerste echte kennismaking met de wereld van de predictiemodellen. Eindeloze discussies over items, sub-items, applicability en adherence later (en die verliepen ook wel eens per mail en zelfs via handgeschreven krabbels), weet ik er heel wat meer van af en het leverde een paar fraaie hoofdstukken op voor in dit proefschrift.

Eigenschappen die jou typeren zijn je enthousiasme, daadkracht en grenzeloze optimisme. Die zorgden telkens weer voor een positief effect op mijn motivatie en vertrouwen. Met mijn optimisme zit het trouwens ook wel goed, soms heb ik er zelfs iets te veel van. Bedankt voor je geduld als het weer eens langer duurde dan gedacht voor ik iets goed genoeg vond om aan je voor te leggen.

Dr. L. Hooft, beste Lotty, na de verhuizing van Cochrane Netherlands naar Utrecht gingen we intensiever samenwerken, helemaal toen jij mijn copromotor werd. Jouw inbreng is in elk hoofdstuk van dit proefschrift terug te zien. Ik heb genoten van alle brainstorms en discussies. Je bent scherp, creatief en net zo van de details als ik (heel gevaarlijk, maar gelukkig zijn we ons ervan bewust). Bovendien had je het ook snel door als het even wat minder ging. Dat je me het vertrouwen gaf om in Genève les te geven bij de WHO, heeft de docent in mij enorm vooruit geholpen. Daarnaast heb ik er een onvergetelijke verjaardag gevierd. Het is verbazingwekkend hoe vaak we hetzelfde denken en we kunnen aan een half woord of blik genoeg hebben. Misschien komt het doordat we de dingen met dezelfde Noord-Hollandse nuchterheid benaderen? Ik kijk er in ieder geval naar uit om samen aan nog heel veel mooie projecten te werken.

Dr. J. B. Reitsma, beste Hans, toen ik bij het Dutch Cochrane Centre kwam werken, had jij het AMC net verruild voor het Julius Centrum. Als bezitter van jouw oude telefoontoestel kende ik je al gauw van naam door alle mensen die nog voor je belden. Toen kon ik nog niet vermoeden dat je een aantal jaren later mijn copromotor zou worden. Ik wil je bedanken voor al je waardevolle suggesties en adviezen. Het was fijn dat jij als relatieve buitenstaander meedacht over de deïmplementatiereviews. Je bent altijd bereid om ergens nog eens een keer rustig naar te kijken. Vanuit een eerste, vage schets op een kladblaadje kan dan zomaar een verhelderende figuur voor in het artikel ontstaan. Hoewel jouw kritische vragen niet zelden extra werk opleveren, wordt het eindresultaat er altijd beter van. Bedankt ook voor de interesse die je altijd toont in mijn persoonlijke welzijn.

Dank aan alle coauteurs voor hun inhoudelijke bijdragen en aan Mariska en Monique voor de ondersteuning bij het versturen van de enquêtes: zonder jullie waren de hoofdstukken er niet gekomen.

Dear Romin, we spent quite some time discussing prediction model studies and extracting data for the TRIPOD adherence project. I really enjoyed our collaboration and would like to thank you for all your hard work and friendship.

Ondanks alle uitdagingen en de enorme hoeveelheid studies om door te ploegen, ben ik blij dat het 'Doen of laten?'-programma op mijn pad kwam. Een relatief onbekend

onderwerp, de directe relatie met de klinische praktijk en de samenwerking met vele partijen maakten het een leerzaam en interessant project. Christiana, Tijn, Simone, Jan-Willem en vooral Eva, bedankt dat we samen hierin konden optrekken.
Dear Toshi, I am very lucky to have you as a colleague. Thank you so much for your efforts and all our valuable discussions.
Claudia, onvermoeibaar offerde je zelfs je vrije tijd op om aan onze review te werken; ik ben buitengewoon dankbaar voor al je hulp.

De leden van de beoordelingscommissie en promotiecommissie, prof. dr. L. Schoonhoven, prof. dr. D. van der Windt, prof. dr. P.W.B. Nanayakkara, prof. dr. M.J. Schuurmans, prof. dr. N.J. de Wit, prof. dr. J.H.H. van de Wijgert en prof. dr. W.A. van Klei, wil ik hartelijk bedanken voor hun bereidheid mijn proefschrift te beoordelen en op 30 juni oppositie te voeren.

Ik ben blij met alle collega's die ik om me heen heb (gehad). Zij hebben ervoor gezorgd dat ik me thuis voelde en boden me zo de ruimte om me te kunnen ontwikkelen.

Mijn Cochrane Netherlands collega's, René, Anneke, Michiel, Kevin en de 'affiliated researchers', het is fijn om als team met jullie aan allerlei onderzoeks- en onderwijsprojecten te werken. Bedankt voor de gezelligheid, niet alleen op de werkvloer, maar ook tijdens borrels, etentjes en congresbezoeken.
René, je bent altijd wel in voor een praatje. Onze gesprekken beperken zich niet tot het zoeken van literatuur, maar kunnen ook over handige software tools, ideeën voor onderzoek of de serieuzere zaken van het leven gaan. Dankjewel voor je betrokkenheid.
Lieve Anneke, we begonnen op hetzelfde moment in het UMCU en hebben inmiddels aan heel wat projecten samengewerkt. Daarbij vullen we elkaar goed aan en het is prettig dat jij ook zo kritisch bent. Tijdens koffiemomenten en lunchwandelingen kunnen we het over van alles en nog wat hebben en we zijn zelfs samen op reis geweest door Zuid-Afrika. Ik hoop dat we nog een hele tijd collega's zullen blijven. Twee jaar geleden was ik jouw paranimf, een hele eer! Ik ben heel blij dat jij die rol ook op je wilt nemen bij mijn promotie.

De collega's van het Julius Centrum, in het bijzonder het Epi Methoden Team en de STARTblok-collega's, wil ik bedanken voor alle hulp en voor de plezierige sfeer.
Coby, een welgemeend goedemorgen van jou is altijd een prima start van de werkdag!
Alle (oud-)kamergenoten van 6.104 (de gezelligste kamer van het Stratenum, mét uitzicht op de Dom!), bedankt dat ik met jullie de dagelijkse beslommeringen, verwonderingen en frustraties kon delen.

De oud-collega's van het Dutch Cochrane Centre, Hanni, Miranda, Annefloor, Roy, Sharon en Fleur, wil ik bedanken voor de fijne samenwerking.
Lieve Fleur, wij werden tegelijk aangenomen en vormden als kamergenootjes een perfect team: naast ontzettend hard werken en serieuze gesprekken, hebben we ook heel erg veel gelachen. Ik hoop dat jij tijdens de verdediging van dit proefschrift je droom aan het verwezenlijken bent en een prachtige wereldreis maakt.

Mijn oud-collega's van de poli oogheelkunde in Alkmaar mogen niet ontbreken in dit dankwoord. Jullie belangstelling, steun en de bereidheid om een stapje extra te doen voor de dataverzameling voor mijn masterthesis zijn heel belangrijk geweest. Ik denk met ontzettend veel plezier terug aan mijn tijd als optometrist en orthoptist en hoop dat ik jullie zo af en toe blijf tegenkomen.

Ook buiten het werk heb ik het geluk allerlei lieve en leuke mensen om mij heen te hebben, die er altijd voor me zijn, terwijl ik de afgelopen jaren soms te weinig tijd en aandacht voor hen had.

Muziek en sport zijn voor mij belangrijke uitlaatkleppen, die een hoop gezelligheid met zich meebrengen. Dank daarom aan al mijn medezangers en -zwemmers.
Joyce en Marina, ook buiten het zwembad kunnen we goed met elkaar opschieten. Bedankt voor de vele goede gesprekken bij een kop thee of een glas wijn.

Lieve Sieta, onze vriendschap gaat al terug tot op de kleuterschool. Tot en met 6 VWO hebben we in de klas vrijwel altijd naast elkaar gezeten (vrijwel, want niet alle leraren konden ons geklets waarderen) en daarnaast hebben we vele uren samen in het zwembad doorgebracht. Ik bewonder je doorzettingsvermogen en ben blij om te zien dat jij in je carrière bereikt wat je voor ogen had.

Lieve opto-meiden, Henrike, Maartje, Bianca en Wencke, op de opleiding optometrie konden we het al snel heel goed met elkaar vinden. De traditie om twee keer per jaar een dagje samen op stap te gaan, houden we nog steeds in stand. Hoewel we allang niet meer allemaal als optometrist werken, blijven we betrokken bij elkaar; ik ben er blij mee!
Lieve Maartje, de dagjes met jou, Marc en jullie meiden zijn een soort mini-vakanties, vol gezelligheid en goede gesprekken. Het zou leuk zijn als jullie nog eens een echt zonnige zomerdag op het Heemskerkse strand meemaken.

Lieve Henrike, vanaf de allereerste minuut op de Hogeschool Utrecht zijn we onafscheidelijk. Ook bij de OVN trokken we intensief met elkaar op en hebben we van alles beleefd. Ik heb het getroffen met jou als mijn beste vriendin! Al die jaren en diverse vakanties verder zijn we nog steeds niet uitgepraat. Ik kan me niet voorstellen dat dat ooit zal veranderen.

Lieve Marjan, voor een schaatscarrière bleken we niet in de wieg gelegd, maar op de ijsbaan in Alkmaar begon wél een hele dierbare vriendschap. Van schaatsmaatje werd je mijn buurvrouw, wat was dat gezellig! Vaste prik op de zondagavond: samen eten en BZV kijken. Des te leuker dat jij je eigen boer vond! Bedankt voor je vriendschap. Ik voel me altijd heel welkom bij jullie op de boerderij.

Familie zoek je niet uit, die krijg je cadeau; in mijn geval een mooi cadeau! Dank aan alle ooms, tantes, neven en nichten die door de jaren heen steeds weer belangstelling toonden voor mijn opleidingen, werk en de voortgang van dit proefschrift. Lydia en Yvonne, afstand zegt niets, dat is wel gebleken. De band die op de vele zondagochtenden bij oma is ontstaan, is niet kapot te krijgen.

Lieve Inge, zonder jouw belangstelling voor de master EBP was dit proefschrift er nooit gekomen. Ik ging voor de gezelligheid met je mee naar de informatiebijeenkomst, werd enthousiast en uiteindelijk hebben we samen de opleiding doorlopen – hoe bijzonder! Je bent een geweldige zus, bij wie ik altijd terecht kan en op wie ik enorm trots ben. Ik vind het mooi om te zien hoe jij bij Sander je geluk hebt gevonden. Dat je daarvoor naar de andere kant van het land verhuisde, is jammer voor mij, maar jullie zijn een prachtig paar en ik gun jullie alle liefde en geluk samen. Bovendien levert het mij een leuke bestemming op voor een weekendje weg (en daar heb ik nu alle tijd voor!). Natuurlijk ben jij op 30 juni mijn paranimf!

Ten slotte, lest best, mijn lieve ouders. Jullie aandeel in waar ik nu sta, is enorm. Dank jullie wel dat jullie me altijd mijn eigen keuzes lieten maken en mij daarbij onvoorwaardelijk steunden. Ook jullie hulp in praktische zin waardeer ik zeer, ook al ben ik met mijn motto ‘zelf doen’ niet de makkelijkste als het op accepteren van hulp aankomt. Ik ben blij dat jullie in goede gezondheid meemaken dat ik dit promotietraject tot een succesvol einde breng. En ik zeg het veel te weinig, maar nu staat het voor altijd op papier: ik hou van jullie.

## Curriculum vitae

Pauline Heus was born in Heemskerk, the Netherlands, on November 28, 1978. After completing pre-university education at the Augustinus College in Beverwijk in 1997, she enrolled in the bachelor programs of Optometry and Orthoptics at the University of Applied Sciences in Utrecht and graduated in 2001. She worked as an optometrist and orthoptist at Van Els Optiek in Heemskerk (until 2006) and at the ophthalmology department of Northwest Clinics in Alkmaar (until 2014). Between 2006 and 2014, Pauline actively contributed to the profession of optometry in the Netherlands as a board member of the Dutch Optometric Association (Optometristen Vereniging Nederland, OVN) and as an editor of the Dutch optometric journal 'Visus'. From 2010 to 2012 she was a board member of the Quality Register of Allied Health Professions (Kwaliteitsregister Paramedici).

After obtaining a master's degree in Evidence Based Practice (with distinction) at the University of Amsterdam in 2010, Pauline joined Cochrane Netherlands in 2011, at the time based in Amsterdam. Since then, she has conducted various systematic reviews and contributed to evidence syntheses for the National Health Care Institute (Zorginstituut Nederland), the Belgian Health Care Knowledge Centre, World Health Organization, and several evidence-based guidelines. In addition, Pauline is involved in teaching and she provides methodological support to systematic review authors. Pauline is member of the Cochrane Bias Methods Group, Cochrane Knowledge Translation Working Group 'Growing Capacity in Users', and the international GRADE Working Group.

In 2014, Cochrane Netherlands moved to the Julius Center for Health Sciences and Primary Care in Utrecht. In July 2015, Pauline started a PhD trajectory under the supervision of prof. dr. K.G.M. Moons, prof. dr. R.J.P.M. Scholten, dr. L. Hooft and dr. J.B. Reitsma, which resulted in this thesis. During this period she was also involved in the Dutch national program 'To do or not to do?' addressing de-implementation of low-value care.

After obtaining her PhD degree, Pauline will continue her work at Cochrane Netherlands and the Julius Center for Health Sciences and Primary Care.

## List of publications

### Peer reviewed publications

**Heus P**, Reitsma JB, Collins GS, Damen JAAG, Scholten RJPM, et al. Transparent Reporting of Multivariable Prediction Models in Journal and Conference Abstracts: TRIPOD for Abstracts. *Ann Intern Med*. 2020;173(x):xxx-xxx.*[article in press]*

**Heus P**, Damen JAAG, Pajouheshnia R, Scholten RJPM, Reitsma JB, et al. Uniformity in measuring adherence to reporting guidelines: the example of TRIPOD for assessing completeness of reporting of prediction model studies. *BMJ Open*. 2019;9(4):e025611.

Damen JA, Pajouheshnia R, **Heus P**, Moons KGM, Reitsma JB, et al. Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis. *BMC Med*. 2019;17(1):109.

Wei D, **Heus P**, van de Wetering FT, van Tienhoven G, Verleye L, Scholten RJPM. Probiotics for the prevention or treatment of chemotherapy- or radiotherapy-related diarrhoea in people with cancer. *Cochrane Database Syst Rev*. 2018;2018(8).

**Heus P**, Verbeek JH, Tikka C. Optical correction of refractive error for preventing and treating eye symptoms in computer users. *Cochrane Database Syst Rev*. 2018;4.

**Heus P**, Damen JAAG, Pajouheshnia R, Scholten RJPM, Reitsma JB, et al. Poor reporting of multivariable prediction model studies. *BMC Med*. 2018;16(1):120.

Santesso N, Carrasco-Labra A, Langendam M, Brignardello-Petersen R, Mustafa RA, **et al**. Improving GRADE evidence tables part 3. *J Clin Epidemiol*. 2016;74:28-39.

Langendam M, Carrasco-Labra A, Santesso N, Mustafa RA, Brignardello-Petersen R, **et al**. Improving GRADE evidence tables part 2. *J Clin Epidemiol*. 2016;74:19-27.

Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, **et al**. Prediction models for cardiovascular disease risk in the general population. *BMJ*. 2016;353.

Ooft ML, Braunius WW, **Heus P**, Stegeman I, van Diest PJ, et al. Prognostic significance of the EGFR pathway in nasopharyngeal carcinoma. *Biomark Med*. 2015;9(10):997-1010.

Lawrie TA, Winter-Roach BA, **Heus P**, Kitchener HC. Adjuvant (post-surgery) chemotherapy for early stage epithelial ovarian cancer. *Cochrane Database Syst Rev*. 2015;2015(12).

van Hassel D, Coops A, Batenburg R, **Heus P**. Een meting van het substitutiepotentieel binnen de oogzorg. *TSG*. 2013;91(8):497-504.

**Heus P**, Van Trijffel E, Busch EM, Lucas C. Mydriatic visual acuity in diabetic patients. *Optom Vis Sci*. 2013;90(3):249-256.

**Reports and guidelines**

JGZ-richtlijn Opsporen Oogafwijkingen. Nederlands Centrum Jeugdgezondheid; 2019. https://www.ncj.nl/richtlijnen/alle-richtlijnen/richtlijn/opsporen-oogafwijkingen-2019. Accessed 2020.

**Heus P**, Spijker R, Oerbekke M, Hooft L. *Identification and critical appraisal of clinical guidelines for cardiac disease, diabetes and stroke. (Guideline analysis for WHO)*. Cochrane Netherlands; 2019.

Richtlijn Liesbreuk bij volwassenen. Richtlijnendatabase Kennisinstituut Federatie Medisch Specialisten; 2019. https://richtlijnendatabase.nl/richtlijn/liesbreuk_bij_volwassenen/startpagina.html. Accessed 2020.

van Dulmen S, **Heus P**, Kool T, Verkerk E. *Doen of laten in de gezondheidszorg. Een onderzoek naar de mogelijkheden van terugdringen van niet-gepaste zorg*. Nijmegen: IQ healthcare; 2019.

**Heus P**, Doorn Sv, Damen JAAG, Spijker R, Hooft L, Scholten RJPM. *Diagnostiek bij patiënten met slaperigheidsklachten en verdenking op slaapapneu (OSA). Deel 2. Systematische reviews*. Utrecht: Cochrane Netherlands; 2018.

**Heus P**, Damen JAAGJ, E., Spijker R, Scholten RJPM. *Diagnostiek bij patiënten met slaperigheidsklachten en verdenking op slaapapneu (OSA). Deel 1. Analyse richtlijnen*. Utrecht: Cochrane Netherlands; 2018.

Richtlijn Lokale anesthesie in de oogheelkunde. Richtlijnendatabase Kennisinstituut Federatie Medisch Specialisten; 2017. https://richtlijnendatabase.nl/richtlijn/lokale_

anesthesie_in_de_oogheelkunde/preoperatief_beleid_anesthesie_oogheelkunde.html

Vergote I, Vlayen J, **Heus P**, Hoogendam JP, Damen JAAG, et al. *Ovarian cancer: diagnosis, treatment and follow-up. Good Clinical Practice (GCP).* Brussels: Belgian Health Care Knowledge Centre (KCE) 2016. KCE Reports 268. D/2016/10.273/49.

van Doorn S, Boonacker CWB, de Groot JAH, **Heus P**, Naaktgeboren CA, et al. *Risicostratificatie en Diagnostiek bij patiënten met klachten van pijn op de borst verdacht voor coronaire hartziekte. Deel 1. Analyse richtlijnen.* Utrecht: Cochrane Netherlands; 2016.

Boonacker CWB, Doorn Sv, de Groot JAH, **Heus P**, Hooft L, et al. *Risicostratificatie en Diagnostiek bij patiënten met klachten van pijn op de borst verdacht voor coronaire hartziekte. Deel 2. Systematische reviews.* Utrecht: Cochrane Netherlands; 2016.

NHG Werkgroep Visusklachten. NHG Standaard Visusklachten (tweede herziening). *Huisarts Wet.* 2015;58(10):532-540.

Grégoire V, Leroy R, **Heus P**, Hooft L, van de Wetering F, et al. *Oropharyngeal, hypopharyngeal and laryngeal cancer: diagnosis, treatment and follow-up. Good Clinical Practice (GCP).* Brussels: Belgian Health Care Knowledge Centre (KCE); 2015. KCE Reports 256. D/2015/10.273/105.

Scholten RJPM, **Heus P**, van de Wetering FT, van Enst WA, Limpens CEJM, Hooft L. *Sleeve Gastrectomie, een systematische review: update.* Amsterdam: Dutch Cochrane Centre; 2013.

Langendam MW, Hooft L, **Heus P**, Scholten RJPM, Bossuyt PMM. *Alternatieven voor Randomized Controlled Trials in onderzoek naar de effectiviteit van interventies. Eindrapport.* Amsterdam: Dutch Cochrane Centre / Afdeling Klinische Epidemiologie, Biostatistiek en Bioinformatica, Academisch Medisch Centrum Amsterdam; 2013.

Hooft L, **Heus P**, van de Wetering FT, van Enst WA, Scholten RJPM. *Systematic review voor het advies van de Gezondheidsraad "Lyme onder de loep".* Den Haag: Gezondheidsraad; 2013. 2013/12.

Grégoire V, Leroy R, **Heus P**, van de Wetering F, Scholten R, et al. *Oral cavity cancer: diagnosis, treatment and follow-up. Good Clinical Practice (GCP).* Brussels: Belgian Health Care Knowledge Centre (KCE); 2013. KCE Reports 227. D/2013/10.273/58.

Verleye L, van de Wetering FT, **Heus P**, Scholten RJPM, Vlayen J. *Ondersteunende therapie bij kanker - Deel 2: preventie en behandeling van nevenwerkingen van chemotherapie en radiotherapie. Good Clinical Practice (GCP).* Brussel: Federaal Kenniscentrum voor de Gezondheidszorg (KCE); 2012. KCE Reports 191A. D/2012/10.273/86.

Langendam MW, Hooft L, **Heus P**, Van de Glind EMM, van Enst WA, et al. *Medicamenteuze en niet-medicamenteuze interventies voor patiënten met dementie, een scoping review.* Amsterdam: Dutch Cochrane Centre; 2012.

Hooft L, Van de Glind EMM, Langendam MW, Brexkens B, **Heus P**, Van Enst WA. *De onzichtbare kant van Parkinson. Inventarisatie van de beschikbare literatuur over afnemende cognitieve vaardigheden bij de Ziekte van Parkinson in het beginstadium.* Amsterdam: Dutch Cochrane Centre; 2012.

Langendam MW, **Heus P**, van de Wetering FT. *ACE-remmers en AT1-antagonisten bij nefropathie, nierinsufficiëntie en hartfalen: update van de relevante richtlijnen.* Amsterdam: Dutch Cochrane Centre; 2011.

Elbers GMH, van Enst WA, **Heus P**, Kramer SF, Van de Wetering FT, et al. *Oefentherapie bij 'Chronisch Obstructieve Longziekten' (COPD).* Amsterdam: Dutch Cochrane Centre; 2011.

**Conference presentations**

**Heus P**, Reitsma JB, Collins GS, Damen JAAG, Scholten RJPM, et al. Transparent reporting of multivariable prediction models in journal and conference abstracts: TRIPOD for Abstracts. Poster presentation at REWARD | EQUATOR Conference 2020; Berlin, Germany.

**Heus P**, Kruithof E, Damen JAAG, Verhoef-Jurgens MS, Scholten RJPM, et al. Implementation of reporting guidelines: a survey of editors and 'Instructions to authors' of medical journals. Poster presentation at REWARD | EQUATOR Conference 2020; Berlin, Germany.

**Heus P**, Damen JAAG, Lamberink HJ, Otte WM, Tijdink JK, et al. Responsible research: trends over time in randomized clinical trials. Oral presentation at Dutch Epidemiology Congress WEON 2019; Groningen, the Netherlands.

**Heus P**, van Dulmen SA, Weenink JW, Naaktgeboren CA, Verkerk EW, et al. Effectiveness of strategies to reduce low-value care: a systematic review of de-implementation studies. Poster presentation at Cochrane Colloquium 2018; Edinburgh, United Kingdom.

**Heus P**, Naaktgeboren C, van Dulmen S, Weenink JW, Spijker R, et al. Determinants for successful de-implementation of low-value services: a systematic review. Oral presentation at Cochrane Colloquium 2016; Seoul, Republic of Korea.

**Heus P**, Hooft L, Reitsma JB, Scholten RJ, Altman DG, et al. Reporting of clinical prediction model studies in journal and conference abstracts: TRIPOD for Abstracts. Oral presentation at Cochrane Colloquium 2016; Seoul, Republic of Korea.

**Heus P**, Damen JAAG, Pajouheshnia R, Scholten RJ, Reitsma JB, et al. More than half of the TRIPOD items are inadequately reported in prediction modelling studies. Poster presentation at Methods for Evaluating Medical Tests and Biomarkers Symposium 2016; Birmingham, United Kingdom.

**Heus P**, Damen J, Scholten R, Reitsma JB, Collins GS, et al. Quality of reporting of clinical prediction model studies - adherence to TRIPOD. Oral presentation at Cochrane Colloquium 2015; Vienna, Austria.

**Heus P**, Van Trijffel E, Busch EM, Lucas C. Visusbepaling in mydriasis: een gerandomiseerde equivalentiestudie. Oral presentation at Jaarvergadering Nederlands Oogheelkundig Gezelschap 2011; Groningen, the Netherlands.

**Heus P**, Van Trijffel E, Busch EM, Lucas C. Randomized controlled equivalence trial to investigate the impact of determining VA prior to or after pupillary dilation on the optimum interval between visits for diabetic patients. Poster presentation at European Academy of Optometry and Optics 2011; Prague, Czech Republic.