Check for updates

When and how to use data from randomised trials to develop or validate prognostic models

Romin Pajouheshnia,[1,2] Rolf H H Groenwold,[3] Linda M Peelen,[1] Johannes B Reitsma,[1,4] Karel G M Moons[1,4]

[1]Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht University, 3508 GA Utrecht, Netherlands
[2]Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, Netherlands
[3]Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, Netherlands
[4]Cochrane Netherlands, Utrecht, Netherlands

Correspondence to:
Romin Pajouheshnia
R.Pajouheshnia@uu.nl
(ORCID 0000-0002-4208-3583)

Cite this as: *BMJ* 2019;365:l2154
http://dx.doi.org/10.1136/bmj.l2154

Accepted: 26 March 2019

Prediction models have become an integral part of clinical practice, providing information for patients and clinicians and providing support for their shared decision making. The development and validation of prognostic prediction models requires substantial volumes of high quality information on relevant predictors and patient health outcomes. Primary data collection dedicated to prognostic model (development or validation) research could come with substantial time and costs and can be seen as a waste of resources if suitable data are already available. Randomised clinical trials are a source of high quality clinical data with a largely untapped potential for use in further research. This article addresses when and how data from a randomised clinical trial can be used additionally for prognostic model research, and provides guidance for researchers with access to trial data to evaluate the suitability of their data for the development and validation of prognostic prediction models.

SUMMARY POINTS

- To minimise research waste, data from randomised clinical trials (RCTs) might be considered for the development or validation of a prognostic prediction model
- Advantages of RCT data can include completeness, quality, detailed protocols, and broad informed consent
- Randomised treatment allocation facilitates the development and validation of prognostic models that predict risk in the presence or absence of a particular treatment
- RCT data might be less suitable because of selective patient or centre inclusion, extraneous trial effects, or overly specialised predictor measurement, which all could limit the generalisability and thus applicability of prognostic models to real life practice
- Other limitations might be surrogate outcomes that are too short term or clinically irrelevant, or an insufficient sample size for prognostic model development or validation
- This paper provides guidance to appraise the suitability of RCT data for prognostic model research by examining both potential benefits and limitations

Prognostic prediction models—or prognostic models—are used to provide probabilistic predictions of an individual's prognosis, which can be used to support patient counselling and evidence based decision making in clinical practice, as well as research.[1] The development and validation of these models requires substantial amounts of high quality patient and clinical data (information on the development and validation of prognostic models can be found elsewhere[2 3]). Although prospective data collection designed specifically to develop or validate a prognostic model is typically advocated,[1] this is often not feasible or desirable due to the costs involved.

Randomised clinical trials (RCTs) provide a tempting alternative data source for the development and validation of prognostic models: in the year 2018, nearly 25 000 RCTs of treatment interventions were published, generating a large quantity of data (see supplement for the search query). Yet the valuable information gathered in RCTs remains largely untapped by the research community, and could be seen as a source of research waste. At the same time, despite the widespread belief that RCTs are the so-called gold standard for data generation, their suitability for answering questions of a descriptive (that is, predictive) nature has been questioned.[1 4 5] This article starts from the perspective that we would like to develop or externally validate a prognostic model and we have access to individual participant data (referred to as "data" in this article) from a relevant phase III RCT. We present the opportunities that RCT data can offer, describe potential limitations that must be considered, and navigate the do's and don'ts of developing or externally validating a prognostic model with RCT data.

Opportunities arising from RCT data use

So far, several prognostic models have been effectively developed and validated using RCT data (table 1). Data generated by an RCT can confer specific benefits over data from alternative sources, such as from predesigned observational studies, electronic health records, disease specific registers, or administrative medical databases. We outline the key opportunities that RCT data might provide when developing or externally validating a prognostic model.

Data quality: completeness

Missing data are a serious and almost ubiquitous issue for studies that develop or externally validate

BMJ: first published as 10.1136/bmj.l2154 on 29 May 2019. Downloaded from http://www.bmj.com/ on 10 February 2020 at Utrecht University Library. Protected by copyright.

prognostic models.[18 19] To develop a prognostic model, one ideally needs complete information on all candidate predictors and outcomes, measured in all individuals in the study. External validation requires complete information on all predictors and outcomes of the model that is under evaluation. Information on predictor variables that are not routinely measured in practice could have limited functional value, as discussed later. But for variables that can be readily measured in practice, complete recording of measurements can reduce the risk of bias (due to selective "missingness" of data) and improve the certainty around model coefficients in a development study or measures of model performance in an external validation study. Although many methods can handle missing data, the best solution is undoubtedly its prevention.

The completeness of data from RCTs can be an important asset for prognostic model studies. Throughout the design and conduct of an RCT, several strategies can be used to help collect a complete set of information on predictors and outcomes in all trial participants.[20] These methods might include the training of research staff before starting data collection, and incentives for data collectors to collect complete information. While these strategies might be a challenge in multicentre trials, a common shared protocol can be established to help maintain consistent data collection across trial centres. A unique feature of RCTs is detailed study monitoring, usually by several separate committees.[21] Trial oversight committees, such as data monitoring committees, monitor the presence of missing data in a trial. These efforts work synergistically with central and onsite monitoring to keep track of missing data, which can help to identify and prevent additional missing data.

In addition, RCT data can include detailed information on important post-baseline events, which could affect the prognosis of participants. Such details are often not available for all patients in observational databases. Post-baseline events (such as changes in treatment, the use of rescue drug treatments, or competing outcome events) might need to be accounted for when developing or externally validating a prognostic model, and should be reported alongside the results.[22 23]

Data quality: accuracy and consistency

Accurate and consistent predictor and outcome information is a requisite for accurate prognostic models. The accuracy of predictor measurements in a prognostic model study should reflect the accuracy of those measurements in clinical practice, as discussed later. However, concerns have been raised over the quality of the recording of information that is collected from patients in routine practice. RCTs are commonly regarded as a source of high quality health data. As with data completeness, considerable amounts of time and money are spent to ensure that data are correctly measured and recorded.

Firstly, adherence to the trial protocol and standard operating procedures facilitates the accurate and consistent measurement of predictors and outcomes, in particular for specific variables of interest in the trial (although this might not reflect actual variation in practice, as discussed later). Secondly, case report forms require the recording of detailed patient and clinical information and can help to prevent the recording of impossible values, forming a part of the quality assurance process in an RCT.[24] Thirdly, as with data completeness, study monitoring in RCTs helps to maintain accuracy and consistency in the recorded

Table 1 | Examples of studies in which prognostic models have been developed or validated using data from randomised clinical trials (RCTs)

Prognostic model	Clinical use	Data source	Study type†	Sample size*
IMPACT model[6]	Risk of outcome after traumatic brain injury	8 RCTs, 3 observational studies	Development	8509 (5748)
TIMI risk score[7]	Risk of death or ischaemic events in patients with unstable angina/non-ST segment elevation myocardial infarction	2 RCTs (TIMI 11B, ESSENCE)	Development	1957 (327)
EORTC risk tables[8]	Prediction of recurrence of stage Ta T1 bladder cancer	7 RCTs (EORTC trials)	Development	2596 (1240)
ADVANCE cardiovascular risk model[9]	4 year cardiovascular disease risk in patients with type II diabetes	1 RCT (ADVANCE)	Development	7168 (473)
S_2TOP-BLEED[10]	Risk of major bleeding in patients with a transient ischaemic attack/ischaemic stroke on antiplatelet agents	6 RCTs (CAPRIE, ESPS-2, MATCH, CHARISMA, ESPRIT, PRoFESS)	Development	43 112 (1530)
Neonatal metabolic acidosis models[11]	Risk of neonatal acidosis at birth	1 RCT	Validation	5049 (54)
MCL International Prognostic Index[12]	Prediction of mantel cell leukaemia survival	2 RCTs (MCL Younger, MCL Elderly)	Validation	958 (316)
EFFECT model[13]	Risk of mortality within one year of hospital admission for heart failure	1 RCT (EVEREST)	Validation	2662 (712)
SYNTAX score II model[14]	Mortality prediction after percutaneous coronary intervention or coronary artery bypass graft	2 RCTs (BEST, PRECOMBAT)	Validation	1480 (90)
OHTS-EGPS model[15]	5 year risk of open angle glaucoma	2 RCTs, 2 observational studies	Validation	1038 (105)

*Number of participants (number of events).
†Development means the derivation of a prognostic model by selecting predictors and combining them statistically into a multivariable model, with or without internal validation[16 17] (for reference 17, see figure 3 (types 1a, 1b, 2a, and 2b)); validation means the evaluation of the performance of a prognostic model in a separate, external dataset[3 17] (for reference 17, see figure 3 (types 3 and 4)).

BMJ: first published as 10.1136/bmj.l2154 on 29 May 2019. Downloaded from http://www.bmj.com/ on 10 February 2020 at Utrecht University Library. Protected by copyright.

BMJ: first published as 10.1136/bmj.l2154 on 29 May 2019. Downloaded from http://www.bmj.com/ on 10 February 2020 at Utrecht University Library. Protected by copyright.

data. For example, central monitoring includes the checking of data for unusual patterns or implausible values.[21] In addition, source data verification and electronic data capture methods form an additional layer of data validation.[25] Finally, a centralised system for the adjudication of outcome events can be especially important when outcome measurements are subjective. Altogether, these systems and processes can yield data that satisfy the quality requirements of prognostic model studies.

Protocol and records

A trial protocol provides information on the modality and timing of predictor and outcome measurements. Firstly, the protocol promotes the standardisation of measurements and the recording of any protocol deviations. As discussed earlier, this standardisation can improve the accuracy and consistency in the recording of measurements, but could lead to issues with the generalisability of a model if deviations in predictor measurement are flagged and corrected (as discussed later). In addition, the details recorded in a protocol might provide insight into the suitability of a predictor for inclusion in a prognostic model. For example, if the protocol states that a certain predictor should be measured at a time point that is not relevant to routine clinical practice, one might not select the predictor.

Secondly, knowledge of the operationalisation of predictor or outcome measurements provides insight into how well a model can perform in practice, and can inform the assessment of the risk of bias when reviewing a prognostic model study.[26] In addition, information on how and when variables were collected and recorded might have predictive value. For example, the timing of measurements (eg, whether taken during the day or night) can be highly predictive of clinical outcomes.[27]

Treatments

Often in practice, healthcare providers are interested in asking: "What will happen to the patient if I do not treat them?" With this question, prognostic models can be used to support clinical decisions as well as provide information to healthcare providers and patients for counselling.[1] For this purpose, prognostic models must predict risks for patients in the absence of a certain treatment—which can prove challenging in non-RCT data, because of the non-random use of treatments by patients,[28] and because advanced statistical methods might be needed to correctly account for this.[29 30] In the case of RCT data, the effect of treatment use can be solved by simply developing or validating the prognostic model in the control trial arm (control treatment, untreated, or placebo treated) or by including the randomised treatment as a predictor in the model, along with terms for any other treatment-predictor interactions (model development only).[24] However, the placebo arm of a placebo controlled trial might not represent truly untreated patients in usual practice (as discussed later), whereas the control arm of a randomised pragmatic or comparative effectiveness trial better reflects daily practice.

Threats to the viability of RCT data use

Available data from an RCT can have several limitations that might reduce the viability of its use for prognostic model development and validation. Below, we present key challenges when considering RCT data use for prognostic model research. Where necessary, the issues are discussed separately for model development and model validation.

Consent

Consent to reuse RCT data for prognostic modelling might not have been given by the trial participants. By contrast with data repositories established specifically for scientific research purposes (eg, UK Biobank[31]), which have very broad consent for data reuse,[32] trials might not always have asked for a sufficiently broad consent. However, compared with routinely collected data (which have even greater consent challenges in light of the 2016 EU General Data Protection Regulation[33]), RCT data might prove more accessible, especially if trials begin to adopt broad consent for data reuse, as recommended.[34] It is likely that researchers will need to consult their institutional review board before using RCT data for secondary analysis, but whether this satisfies ethical and legal requirements needs further examination.

Selective inclusion of centres

The centres that participate in RCTs might not be representative of medical practice in general.[35] Specifically, generalisability of a prognostic model or the findings of a validation study might be limited if only specialist trial centres (eg, academic medical centres) or experienced clinicians with high performance ratings were invited to participate.[36] In such cases, the associations between predictors and the outcome, and the incidence of outcomes could be different in the trial setting compared with routine clinical practice, of which both could affect the performance of a prognostic model.[37]

Selective eligibility and enrolment

RCTs commonly have narrow participant eligibility criteria, for example, often excluding patients who are frail, who have multimorbidity, or who are vulnerable.[38-42] At the same time, some of the most challenging clinical decisions are for these groups of patients. Thus, RCTs might not provide sufficient information for prognostic research in these clinically relevant patient subgroups. When developing or validating a prognostic model using a selective subset of patients, the predictor effects and functional forms

of their associations with an outcome are assumed to be the same across the patient subsets included and excluded from the RCT. In addition, the participants invited to enter an RCT and those who actually enrol and remain in the trial until completion can substantially differ.[43] For example, the requirement of informed consent from participants has been shown to result in differences between the patients enrolled and not enrolled in a trial.[44 45] As with selective eligibility, this challenge can limit the value of RCT data for prognostic model development; it might not be as problematic for external validation, but could limit the generalisability of results to broader patient populations.

Predictor measurement

As discussed earlier, the strengths of protocol driven data collection by trained research staff highlight that improved availability, accuracy, and consistency of clinical measurements can improve the viability of a prognostic model.[46] With this opportunity, however, come challenging threats. For the purpose of prognostic prediction, the measurement of predictors should closely reflect how they are measured in regular clinical practice. Thus, the use of unrealistically accurate measurements—which could occur if specialist personal or equipment are used in an RCT—when developing a prognostic model could reduce the generalisability of the model to clinical practice,[47] and the findings of a validation study might not represent how the model will truly perform in practice.[48] In addition, it is essential that any predictors considered when developing a prognostic model are (or potentially will be) routinely measured in practice. Supplemental variables collected in an RCT should not be incorporated in a prognostic model if they will not be available in practice.

Extraneous trial effects

The effects of trial enrolment on participant behaviour have been documented extensively, which can vary greatly between trials.[49] Knowledge of enrolment in a trial can lead to participants behaving differently, even reporting more optimistic outcomes,[50 51] an effect commonly termed the "Hawthorne effect." The enrolment of a centre in an RCT might also affect the behaviour of healthcare professionals and as a result the prognosis of a patient enrolled in a trial might be better than if the patient had received routine care.[52] In placebo controlled trials, patients on placebo do not reflect usual or current care, and might also show a placebo or nocebo effect, which could positively or negatively affect their outcomes.[53] If data from a control arm with a strong placebo effect are used to develop a prognostic model to predict a subjective outcome (such as pain experience), the model might underestimate the outcome when applied in practice.

The protocol effect or care effect can arise when the adherence of centres to a strict trial protocol might improve patient outcomes (eg, through additional monitoring) compared with patients not enrolled in the trial.[54 55] The presence of these effects could hamper the generalisability of models developed or validated using RCT data to clinical practice, possibly due to close monitoring or specialised care being specified in the trial protocol, or because of a subconscious effect that trial participation has on care givers. In both cases, if participation in a trial results in better patient outcomes, a prognostic model developed using these data might underestimate risks when the model is applied in practice. Thus, a trial with strong extraneous effects might not provide suitable data for prognosis research.

Short term and surrogate outcomes

Long term, patient relevant outcomes are often of interest when making prognostic predictions in daily practice. For example, models to predict cardiovascular disease risk are commonly designed to predict outcomes within 10 years.[56] Development and validation of such models require very long follow-up, which is rarely available in RCTs. However, unlike the validation of a model for predicting long term prognosis with short term outcome data (which is not advisable), there could still be medical use of a prognostic model developed with short term outcomes. In addition, RCTs often opt for surrogate endpoints to replace more costly long term outcomes.[57] If a prognostic model is to be used to inform patients and healthcare professionals, surrogate endpoints could have insufficient clinical relevance if the surrogate is imperfectly correlated with the clinical outcome, whether used to develop or validate a prognostic model.

Sample size

Research on the development of prognostic models often requires substantial samples. While no consensus currently exists on the sample sizes required for prognostic prediction, the required sample size depends on several factors, including the number of predictors, total sample size, and number or proportion of events.[58 59] Thus, large sample sizes could be needed to reliably develop a prognostic model, especially when tens or hundreds of candidate predictor variables are considered. Similarly, reliable prognostic model validation requires data samples with a minimum of several hundreds of outcome events.[60] Obviously, RCTs are not designed and powered with prognosis research in mind. Thus, the number of participants in a single phase III RCT might not be sufficient, and the problem worsens if smaller phase IIb RCTs are considered. Although approaches such as penalised regression can help to prevent the overfitting of prognostic models in small datasets,[61] large samples may yet be needed for modern modelling techniques.[62] Data from large, multicentre trials can, however, provide a solution to this issue. In addition, as seen in table 1, the combination of individual participant data from more than one RCT can greatly increase the amount of available data.

BMJ: first published as 10.1136/bmj.l2154 on 29 May 2019. Downloaded from http://www.bmj.com/ on 10 February 2020 at Utrecht University Library. Protected by copyright.

How and when to use trial data for prognostic prediction research

When data from an RCT are available, researchers must weigh the advantages (data quality, completeness, treated and untreated arms, and protocol) against any limitations, both described earlier. We suggest that the decision process be separated as follows:

- Criteria that must be met: there must be acceptable patient consent (or under certain conditions a waiver by an institutional review board [63]) for reuse of the RCT data for prognostic prediction research.
- Criteria that could seriously limit the usefulness of the data: insufficient sample size or follow-up, or no availability of important predictors or outcomes will seriously limit the suitability of RCT data.
- Criteria that could limit the usefulness of the data: selection of patients or centres, experimental effects, and predictor measurements highly driven by the protocol could all limit how representative the trial data are of the target population for the prognostic model.

To aid in this process, table 2 presents a series of questions to ask when assessing whether data from an RCT are suitable for developing or validating a prognostic model. For each situation, general advice is provided to help researchers reach a decision. The decision to use a given set of data from an RCT will depend on the specific research question and remains largely subjective. In addition, to help gain an overall picture of the suitability of an RCT as a whole, researchers can benefit from constructing a diagram, such as in figure 1. In this fictional example, consent for secondary use of the data was available and the data received a high "score" for this criterion, after which the remaining criteria were assessed. From this, we see that the dataset might be a good candidate for a validation study, but centre and participant selection could limit the generalisability of prognostic models developed using the data. With such a picture, researchers can decide whether the benefits of using the RCT data outweigh the limitations. Finally, if a decision is made to use the RCT data, this information can be used when reporting any study limitations.

Use of data from more than one study

As seen in table 1, multiple RCTs can be combined when developing or validating a prognostic model, which has the clear advantage of increasing the number of patients and outcomes in the analysis, and can provide an opportunity to assess the impact of differences in definitions and measurements on model performance. The IMPACT model,[6] for example, combined data from both RCTs and observational studies. Datasets ranged in size from 139 to 1574 patients (8509 in total), providing much more information than any single study. A cross validation procedure was performed to assess performance of the model across the studies, giving more insight into the robustness of the findings. How data from RCTs and observational studies should be best combined for prognostic model studies requires further research. For now, we suggest that researchers use figure 1 to assess and compare the suitability of multiple RCTs, and that readers should refer to existing guidance on individual patient data meta-analysis for prognostic modelling studies.[64]

Conclusion

When data from an RCT are available for the secondary purpose of developing or validating a prognostic model, the opportunities and limitations of these data require careful consideration. Available data from RCTs can, if used appropriately, be a viable substitute for costly and labour intensive data collection for prognostic prediction research. By recognising the opportunities

Table 2 | How to assess whether data from a randomised clinical trial is suitable for developing or validating a prognostic model

Consideration	Questions to consider	How to proceed*
Consent	Did patients give consent or has consent been waived for the data to be reused for prognostic prediction research?	If consent for reuse is inadequate, data should not be used.
Selective inclusion of centres	Are the trial centres (their expertise, facilities, and use of complex interventions) representative of centres where you might use the prognostic model?	Development: might not be suitable if trial centres do not represent standard practice. Validation: might still be used, but report limitations to the generalisability of results.
Selective eligibility and enrolment	Did the trial eligibility criteria result in the exclusion of relevant patient groups for the prognostic model? Are patients who did or did not enrol (after invitation), and patients who remained or left the study comparable in terms of their characteristics and possible prognosis?	Development: might not be suitable if participants do not represent the target population for prognostic prediction. Validation: might still be used, but report limitations to the generalisability of results.
Predictor measurement	Were predictors measured as they would be in routine clinical practice?	If the methods of predictor measurement are unrealistic, data might not be suitable.
Extraneous trial effects	Could enrolment in the trial have influenced patient prognosis beyond a treatment effect?	If there is evidence for strong extraneous effects, data might not be suitable.
Short term and surrogate outcomes	Were relevant patient outcomes measured? Is there a sufficiently long follow-up for outcomes?	If outcomes or the timing of their measurement are not relevant, data might not be suitable.
Sample size	How many participants were enrolled and remained in the study? What proportion had the outcome?	Development: consider methods for data with few events,[61] but could be too small for any meaningful modelling. Validation: data might not be suitable.

*If a consideration might seriously limit the usefulness of a randomised clinical trial for prognostic prediction research, suggestions on how to proceed are provided.

BMJ: first published as 10.1136/bmj.l2154 on 29 May 2019. Downloaded from http://www.bmj.com/ on 10 February 2020 at Utrecht University Library. Protected by copyright.

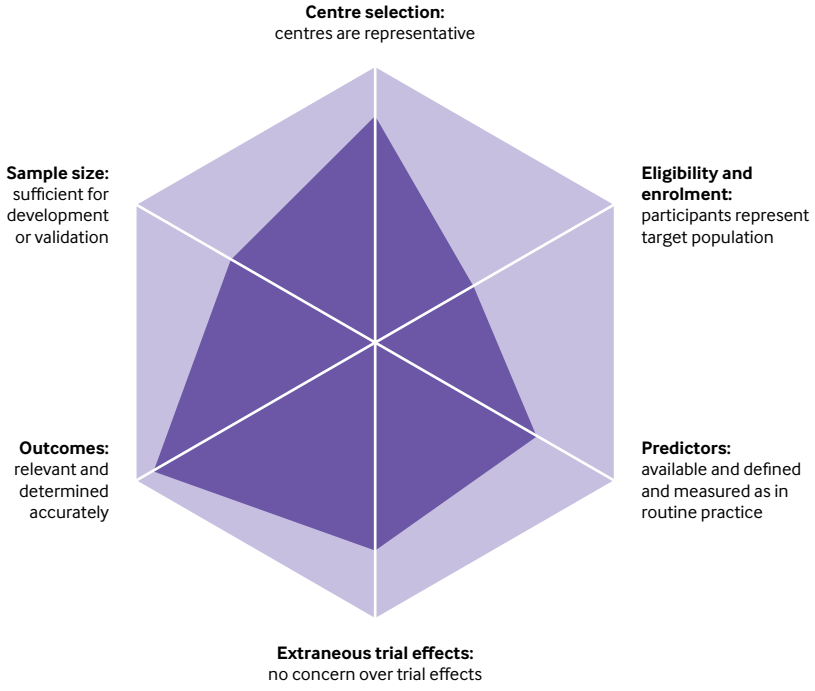

Fig 1 | Graphical example of how suitability of data from a randomised clinical trial can be assessed after consent for use has been confirmed. Assuming consent for use of the data are obtained, researchers might attribute a level of confidence that the RCT does have additional threats (as described earlier). In this example, the dark purple boundaries converge towards the ends of four axes, but limited sample size and representativeness of the participants in the RCT might force researchers to question its suitability for prognostic modelling

that RCT data offer and carefully appraising available data, we can maximise the chance of using data that allow high quality prognostic model research, while avoiding unnecessary, costly primary data collection.

Inevitably, fundamental challenges remain that are universal to the secondary use of data for research, such as the systematic absence of data on certain key predictors. In these circumstances, researchers might consider designing a dedicated study to collect data to develop or externally validate a prognostic model. Alternatively, they could consider greater integration of prognosis research questions during the design of clinical trials, which could help to overcome barriers such as consent. We hope that researchers will cautiously seize the opportunities that data generated by RCTs provide, to improve both the quality and efficiency of future prognostic prediction research.

We thank Rieke van der Graaf for providing insight into ethics and consent for trial data reuse.

Contributors: KGMM conceived the paper objectives. The manuscript was first drafted by RP, which was subsequently reviewed several times by all authors. All authors contributed to the design of the study, manuscript content, and the writing of the manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: There was no specific funding for this paper.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; RG receives funding from the Netherlands Organisation for Scientific Research (project 917.16.430); JBR is supported by a TOP grant from the Netherlands Organisation for Health Research and Development (ZonMw) entitled "Promoting tailored healthcare: improving methods to investigate subgroup effects in treatment response when having multiple individual participant datasets" (grant 91215058); KGMM receives funding from the Netherlands Organisation for Scientific Research (project 9120.8004 and 918.10.615); no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how?*BMJ* 2009;338:b375. doi:10.1136/bmj.b375
2 Steyerberg EW, Moons KG, van der Windt DA, et al, PROGRESS Group. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10:e1001381. doi:10.1371/journal.pmed.1001381
3 Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605. doi:10.1136/bmj.b605
4 Vandenbroucke JP. Observational research, randomised trials, and two views of medical science. *PLoS Med* 2008;5:e67. doi:10.1371/journal.pmed.0050067
5 Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;350:h870. doi:10.1136/bmj.h870
6 Steyerberg EW, Mushkudiani N, Perel P, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med* 2008;5:e165, discussion e165. doi:10.1371/journal.pmed.0050165
7 Antman EM, Cohen M, Bernink PJ, et al. The TIMI risk score for unstable angina/non-ST elevation MI: A method for prognostication and therapeutic decision making. *JAMA* 2000;284:835-42. doi:10.1001/jama.284.7.835
8 Sylvester RJ, van der Meijden AP, Oosterlinck W, et al. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *Eur Urol* 2006;49:466-5, discussion 475-7. doi:10.1016/j.eururo.2005.12.031
9 Kengne AP, Patel A, Marre M, et al, ADVANCE Collaborative Group. Contemporary model for cardiovascular risk prediction in people with type 2 diabetes. *Eur J Cardiovasc Prev Rehabil* 2011;18:393-8. doi:10.1177/1741826710394270

BMJ: first published as 10.1136/bmj.l2154 on 29 May 2019. Downloaded from http://www.bmj.com/ on 10 February 2020 at Utrecht University Library. Protected by copyright.

10 Hilkens NA, Algra A, Diener HC, et al, Cerebrovascular Antiplatelet Trialists' Collaborative Group. Predicting major bleeding in patients with noncardioembolic stroke on antiplatelets: S2TOP-BLEED. *Neurology* 2017;89:936-43. doi:10.1212/WNL.0000000000004289
11 Schuit E, Amer-Wahlin I, Groenwold RH, Mol BW, Moons KG, Kwee A. Prediction of neonatal metabolic acidosis in women with a singleton term pregnancy in cephalic presentation: an external validation study. *Am J Perinatol* 2012;29:681-6. doi:10.1055/s-0032-1314888
12 Hoster E, Klapper W, Hermine O, et al. Confirmation of the mantle-cell lymphoma International Prognostic Index in randomized trials of the European Mantle-Cell Lymphoma Network. *J Clin Oncol* 2014;32:1338-46. doi:10.1200/JCO.2013.52.2466
13 Wessler BS, Ruthazer R, Udelson JE, et al. Regional validation and recalibration of clinical predictive models for patients with acute heart failure. *J Am Heart Assoc* 2017;6:e006121. doi:10.1161/JAHA.117.006121
14 Sotomi Y, Cavalcante R, van Klaveren D, et al. Individual long-term mortality prediction following either coronary stenting or bypass surgery in patients with multivessel and/or unprotected left main disease: an external validation of the SYNTAX score II model in the 1,480 patients of the BEST and PRECOMBAT randomized controlled trials. *JACC Cardiovasc Interv* 2016;9:1564-72. doi:10.1016/j.jcin.2016.04.023
15 Takwoingi Y, Botello AP, Burr JM, et al, Surveillance for Ocular Hypertension Study Group. External validation of the OHTS-EGPS model for predicting the 5-year risk of open-angle glaucoma in ocular hypertensives. *Br J Ophthalmol* 2014;98:309-14. doi:10.1136/bjophthalmol-2013-303622
16 Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ* 2009;338:b604. doi:10.1136/bmj.b604
17 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med* 2015;13:1. doi:10.1186/s12916-014-0241-z
18 Collins GS, Mallett S, Omar O, Yu L-M. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med* 2011;9:103. doi:10.1186/1741-7015-9-103
19 Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014;14:40. doi:10.1186/1471-2288-14-40
20 Little RJ, Cohen ML, Dickersin K, et al. The design and conduct of clinical trials to limit missing data. *Stat Med* 2012;31:3433-43. doi:10.1002/sim.5519
21 Baigent C, Harrell FE, Buyse M, Emberson JR, Altman DG. Ensuring trial validity by data quality assurance and diversification of monitoring methods. *Clin Trials* 2008;5:49-55. doi:10.1177/1740774507087554
22 Pajouheshnia R, Damen JA, Groenwold RH, Moons KG, Peelen LM. Treatment use in prognostic model research: a systematic review of cardiovascular prognostic studies. *Diagnostic and Prognostic Research* 2017;1:15. doi:10.1186/s41512-017-0015-0
23 Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1-73. doi:10.7326/M14-0698
24 Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. *Perspect Clin Res* 2014;5:159-66. doi:10.4103/2229-3485.140555
25 Nahm ML, Pieper CF, Cunningham MM. Quantifying data quality for clinical trials using electronic data capture. *PLoS One* 2008;3:e3049. doi:10.1371/journal.pone.0003049
26 Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;356:i6460. doi:10.1136/bmj.i6460
27 Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study [correction in: *BMJ* 2018;363:k441]. *BMJ* 2018;361:k1479. doi:10.1136/bmj.k1479
28 Groenwold RH, Moons KG, Pajouheshnia R, et al. Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. *J Clin Epidemiol* 2016;78:90-100. doi:10.1016/j.jclinepi.2016.03.017
29 Pajouheshnia R, Peelen LM, Moons KGM, Reitsma JB, Groenwold RHH. Accounting for treatment use when validating a prognostic model: a simulation study. *BMC Med Res Methodol* 2017;17:103. doi:10.1186/s12874-017-0375-8
30 Sperrin M, Martin GP, Pate A, Van Staa T, Peek N, Buchan I. Using marginal structural models to adjust for treatment drop-in when developing clinical prediction models. *Stat Med* 2018;37:4142-54. doi:10.1002/sim.7913
31 Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779. doi:10.1371/journal.pmed.1001779
32 Grady C, Eckstein L, Berkman B, et al. Broad consent for research with biological samples: workshop conclusions. *Am J Bioeth* 2015;15:34-42. doi:10.1080/15265161.2015.1062162
33 EUR-LEX. General data protection regulation. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC (accessed 16 May 2017).
34 Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open* 2017;7:e018647. doi:10.1136/bmjopen-2017-018647
35 Gheorghe A, Roberts TE, Ives JC, Fletcher BR, Calvert M. Centre selection for clinical trials and the generalisability of results: a mixed methods study. *PLoS One* 2013;8:e56560. doi:10.1371/journal.pone.0056560
36 Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet* 2005;365:82-93. doi:10.1016/S0140-6736(04)17670-8
37 Vergouwe Y, Moons KGM, Steyerberg EW. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. *Am J Epidemiol* 2010;172:971-80. doi:10.1093/aje/kwq223
38 Zulman DM, Sussman JB, Chen X, Cigolle CT, Blaum CS, Hayward RA. Examining the evidence: a systematic review of the inclusion and analysis of older adults in randomized controlled trials. *J Gen Intern Med* 2011;26:783-90. doi:10.1007/s11606-010-1629-x
39 Shields KE, Lyerly AD. Exclusion of pregnant women from industry-sponsored clinical trials. *Obstet Gynecol* 2013;122:1077-81. doi:10.1097/AOG.0b013e3182a9ca67
40 Hutchins LF, Unger JM, Crowley JJ, Coltman CAJr, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med* 1999;341:2061-7. doi:10.1056/NEJM199912303412706
41 Lee PY, Alexander KP, Hammill BG, Pasquali SK, Peterson ED. Representation of elderly persons and women in published randomized trials of acute coronary syndromes. *JAMA* 2001;286:708-13. doi:10.1001/jama.286.6.708
42 Schmidt AF, Groenwold RHH, van Delden JJM, et al. Justification of exclusion criteria was underreported in a review of cardiovascular trials. *J Clin Epidemiol* 2014;67:635-44. doi:10.1016/j.jclinepi.2013.12.005
43 van Staa TP, Dyson L, McCann G, et al. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess* 2014;18:1-146. doi:10.3310/hta18430
44 Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Threats to applicability of randomised trials: exclusions and selective participation. *J Health Serv Res Policy* 1999;4:112-21. doi:10.1177/135581969900400210
45 Junghans C, Feder G, Hemingway H, Timmis A, Jones M. Recruiting patients to medical research: double blind randomised trial of "opt-in" versus "opt-out" strategies. *BMJ* 2005;331:940. doi:10.1136/bmj.38583.625613.AE
46 Khudyakov P, Gorfine M, Zucker D, Spiegelman D. The impact of covariate measurement error on risk prediction. *Stat Med* 2015;34:2353-67. doi:10.1002/sim.6498
47 Luijken K, Groenwold RH, van Calster B, Steyerberg EW, van Smeden M. Impact of predictor measurement heterogeneity across settings on performance of prediction models: a measurement error perspective. arXiv:1806.10495 [Preprint]. 2018. https://arxiv.org/abs/1806.10495
48 Pajouheshnia R, van Smeden M, Peelen LM, Groenwold RHH. How variation in predictor measurement affects the discriminative ability and transportability of a prediction model. *J Clin Epidemiol* 2019;105:136-41. doi:10.1016/j.jclinepi.2018.09.001
49 McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. *BMC Med Res Methodol* 2007;7:30. doi:10.1186/1471-2288-7-30
50 McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol* 2014;67:267-77. doi:10.1016/j.jclinepi.2013.08.015
51 Wolfe F, Michaud K. The Hawthorne effect, sponsored trials, and the overestimation of treatment effectiveness. *J Rheumatol* 2010;37:2216-20. doi:10.3899/jrheum.100497
52 Sedgwick P, Greenwood N. Understanding the Hawthorne effect. *BMJ* 2015;351:h4672. doi:10.1136/bmj.h4672
53 Finniss DG, Kaptchuk TJ, Miller F, Benedetti F. Biological, clinical, and ethical advances of placebo effects. *Lancet* 2010;375:686-95. doi:10.1016/S0140-6736(09)61706-2
54 Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993;342:1317-22. doi:10.1016/0140-6736(93)92244-N

BMJ: first published as 10.1136/bmj.l2154 on 29 May 2019. Downloaded from http://www.bmj.com/ on 10 February 2020 at Utrecht University Library. Protected by copyright.

55 Braunholtz DA, Edwards SJL, Lilford RJ. Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect". *J Clin Epidemiol* 2001;54:217-24. doi:10.1016/S0895-4356(00)00305-X
56 Damen JAAG, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* 2016;353:i2416. doi:10.1136/bmj.i2416
57 Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled?*Ann Intern Med* 1996;125:605-13. doi:10.7326/0003-4819-125-7-199610010-00011
58 van Smeden M, Moons KG, de Groot JA, et al. Sample size for binary logistic prediction models: Beyond events per variable criteria. *Stat Methods Med Res* 2018;962280218784726.
59 Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 2019;38:1276-96. doi:10.1002/sim.7992
60 Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;35:214-26. doi:10.1002/sim.6787
61 Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events [correction in: *BMJ* 2016;353:i3235]. *BMJ* 2015;351:h3868. doi:10.1136/bmj.h3868
62 van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol* 2014;14:137. doi:10.1186/1471-2288-14-137
63 Organization WH. *Sciences CfIOoM. International ethical guidelines for health-related research involving humans.* Council for International Organizations of Medical Sciences, 2016.
64 Debray TP, Riley RD, Rovers MM, Reitsma JB, Moons KG, Cochrane IPD Meta-analysis Methods group. Individual participant data (IPD) meta-analyses of diagnostic and prognostic modeling studies: guidance on their use. *PLoS Med* 2015;12:e1001886. doi:10.1371/journal.pmed.1001886

Web appendix: Supplemental material

BMJ: first published as 10.1136/bmj.l2154 on 29 May 2019. Downloaded from http://www.bmj.com/ on 10 February 2020 at Utrecht University Library. Protected by copyright.

No commercial reuse: See rights and reprints http://www.bmj.com/permissions Subscribe: http://www.bmj.com/subscribe