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 Moons1,4

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.

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 elsewhere2 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

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
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. 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. 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.

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. Thirdly, as with data completeness, study monitoring in RCTs helps to maintain accuracy and consistency in the recorded

<table>
<thead>
<tr>
<th>Prognostic model</th>
<th>Clinical use</th>
<th>Data source</th>
<th>Study type†</th>
<th>Sample size*</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMPACT model</td>
<td>Risk of death or ischaemic events in patients with unstable angina/non-ST segment elevation myocardial infarction</td>
<td>2 RCTs (TIMI 11B, ESSENCE)</td>
<td>Development</td>
<td>1957 (327)</td>
</tr>
<tr>
<td>TIMI risk score</td>
<td>Prediction of recurrence of stage Ia II bladder cancer</td>
<td>7 RCTs (EORTC trials)</td>
<td>Development</td>
<td>2596 (1240)</td>
</tr>
<tr>
<td>EORTC risk tables</td>
<td>Prediction of recurrence of stage Ia II bladder cancer</td>
<td>6 RCTs (CAPRIE, ESPRIT, PROFESS)</td>
<td>Development</td>
<td>43112 (1530)</td>
</tr>
<tr>
<td>ADVANCE cardiovascular risk model</td>
<td>4 year cardiovascular disease risk in patients with type II diabetes</td>
<td>1 RCT (ADVANCE)</td>
<td>Development</td>
<td>7168 (473)</td>
</tr>
<tr>
<td>S TOP-BLEED10</td>
<td>Risk of major bleeding in patients with a transient ischaemic attack/ischaemic stroke on antiplatelet agents</td>
<td>2 RCTs (MCL Younger, MCL Elderly)</td>
<td>Validation</td>
<td>958 (316)</td>
</tr>
<tr>
<td>Neonatal metabolic acidosis models12</td>
<td>Risk of neonatal acidosis at birth</td>
<td>1 RCT</td>
<td>Validation</td>
<td>5049 (54)</td>
</tr>
<tr>
<td>MCL International Prognostic Index12</td>
<td>Prediction of mantle cell leukaemia survival</td>
<td>2 RCTs (MCL Younger, MCL Elderly)</td>
<td>Validation</td>
<td>958 (316)</td>
</tr>
<tr>
<td>EFFECT model</td>
<td>Risk of mortality within one year of hospital admission for heart failure</td>
<td>1 RCT (EVEREST)</td>
<td>Validation</td>
<td>2662 (712)</td>
</tr>
<tr>
<td>SYNTAX score II model</td>
<td>Mortality prediction after percutaneous coronary intervention or coronary artery bypass graft</td>
<td>2 RCTs (BEST, PRECOMBAT)</td>
<td>Validation</td>
<td>1480 (90)</td>
</tr>
<tr>
<td>OHTS-EGPS model</td>
<td>5 year risk of open angle glaucoma</td>
<td>2 RCTs, 2 observational studies</td>
<td>Validation</td>
<td>1038 (105)</td>
</tr>
</tbody>
</table>

*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.‡ (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. (for reference 17, see figure 3 (types 3 and 4)).
data. For example, central monitoring includes the checking of data for unusual patterns or implausible values.\textsuperscript{21} In addition, source data verification and electronic data capture methods form an additional layer of data validation.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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,\textsuperscript{28} and because advanced statistical methods might be needed to correctly account for this.\textsuperscript{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).\textsuperscript{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\textsuperscript{31}), which have very broad consent for data reuse,\textsuperscript{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\textsuperscript{33}), RCT data might prove more accessible, especially if trials begin to adopt broad consent for data reuse, as recommended.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{37}
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.\(^5\) 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.\(^44 \ 46\) 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 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.
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) 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 careful consideration.

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, 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.

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

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

*If a consideration might seriously limit the usefulness of a randomised clinical trial for prognostic prediction research, suggestions on how to proceed are provided.
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; KG receives funding from the Netherlands Organisation for Scientific Research (project 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.

The BMJ: first published as 10.1136/bmj.l2154 on 29 May 2019. Downloaded from bmj.com on 30 July 2022 by guest. Protected by copyright.


63 Organization WH. *Sciences CFIOoM. International ethical guidelines for health-related research involving humans*. Council for International Organizations of Medical Sciences, 2016.


**Web appendix:** Supplemental material