







BMJ Open Design aspects for prognostic factor studies

Peggy Sekula ¹, Inga Steinbrenner ¹, Ulla T Schultheiss,^{1,2,3} Neus Valveny,⁴ Paola Rebora,^{5,6} Susan Halabi ⁷, Suzanne M Cadarette,⁸ Richard D Riley ^{9,10}, Gary S Collins ¹¹, Willi Sauerbrei,¹² Mitchell H Gail ¹³ for topic group 5 of the STRATOS initiative

To cite: Sekula P, Steinbrenner I, Schultheiss UT, *et al.* Design aspects for prognostic factor studies. *BMJ Open* 2025;**15**:e095065. doi:10.1136/bmjopen-2024-095065

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<https://doi.org/10.1136/bmjopen-2024-095065>).

Received 14 October 2024
Accepted 04 August 2025

ABSTRACT

Prognostic research is clinically relevant and ultimately facilitates stratified medicine. However, its quality and output are limited. More guidance is needed to improve understanding and thus quality. On behalf of the topic group ‘TG5: study design’ of the STRATOS initiative and for the general readership, this article describes key concepts and issues for prognostic factor studies, a sub-area of prognosis research. After providing a general overview on prognosis research, the article covers aspects such as aims, estimands and designs of prognostic factor studies, highlighting standards and current practice. Focusing on prognostic factor studies that assess a single factor at a time and a binary outcome, this article is complemented by a glossary of terms and a list of general aspects to consider in prognostic factor studies.

INTRODUCTION

Prognosis is the prediction of the probable course of a disease or its outcome and has been a feature of medicine since the time of Hippocrates.¹ Accurate prognostic information is valuable for advising patients about likely outcomes and guiding treatment to improve outcomes.² Since treatments can have serious adverse effects, their use may be justified only in patients with a poor prognosis. Consider, for example, chronic kidney disease (CKD), a complex disease of global importance with an increasing estimated prevalence of 8%–16% in the adult population worldwide.³ CKD prognosis includes the risk of progression to kidney failure (KF), as well as the risk of mortality and morbidity, which are increased even in persons with mild CKD.³ Prognostic instruments to identify CKD patients at increased risk for KF or other serious events could be valuable tools for informing patient counselling, clinical management and healthcare resource planning.⁴

Prognostic information is also useful in the design and analysis of clinical trials.^{2,5} For example, the sample size needed for a clinical trial determining whether a new treatment

reduces the rate of cancer recurrence depends on the prognosis of the trial participants, because the power to detect a reduction in the rate of recurrence depends on the number of observed recurrences.^{6–8} Prognostic variables are also used to stratify participants of clinical studies into risk groups with comparable outcome probabilities across treatment arms to improve the randomisation process. Also, when adjusting for prognostic variables in the actual analysis, the power to detect genuine treatment effects in a clinical trial can be increased.⁹ Furthermore, prognostic variables are often used as candidates for explaining treatment effect heterogeneity (also known as predictive factors¹⁰ or treatment-covariate interactions). Both predictive and prognostic factors facilitate stratified medicine.¹¹

To improve the prediction of outcomes in diseased patients, single prognostic factors are usually combined in a prognostic model (box 1: prognostic model research).¹² Such models typically include demographic factors such as sex and age, and disease-specific factors such as estimated glomerular filtration rate for KF or prostate-specific antigen for prostate cancer relapse.^{13–15} Examples of prognostic models frequently used are the ‘Nottingham Prognostic Index’ for predicting breast cancer survival or the ‘Corticosteroid Randomisation After Significant Head Injury’ (CRASH) models for the outcome of traumatic brain injury.^{16,17}

Even when a prognostic model is available for a specific setting, its ability to predict the outcome accurately is often limited. Consequently, there is a need to identify additional prognostic factors (box 1: prognostic factor research) and subsequently to develop a new or to update an existing prognostic model (figure 1).² Despite some success stories, the impact of prognostic factor research (or biomarker research in general)¹⁸ has been limited, as too many proposed factors fail



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to

Dr Peggy Sekula;
peggy.sekula@uniklinik-freiburg.de

Box 1 What is prognosis research?

Prognosis research focuses on the assessment of clinically relevant outcomes in patients with a specific disease, such as chronic kidney disease (CKD) or prostate cancer.³² Four different types of prognosis research can be distinguished:

1. Descriptive prognosis research

Descriptive prognosis research describes the spectrum of possible outcomes and their impact and burden in patients with a specific disease.³¹ This perspective thus provides an overall description of outcomes in populations defined by particular conditions and healthcare settings and forms the rationale for many subsequent studies, such as studies to identify specific prognostic factors and to quantify risks precisely.

Example: Keith *et al* described the course of CKD patients in a large cohort with respect to clinical outcomes and comorbidities.¹¹¹

2. Prognostic factor research (single factor of interest)

The aim of prognostic factor research (figure 1) is to identify individual factors that enable to predict the patients' future outcome.² Such factors do not need to be causally related with the outcome to be prognostic. Prognostic factors should be distinguished from 'predictive factors' that predict the size of a treatment effect.¹⁰

To identify novel prognostic factors, recent studies increasingly use laboratory techniques that permit high throughput measurements from samples from large populations and that allow measurement of many features on each sample ('Omics' data, eg, genomics data).¹¹²

Identifying novel factors (discovery) usually starts with tests for an association with an outcome, including their characterisation in patients with and without the outcome. After its discovery, further studies are needed to estimate risk-factor specific disease risk, to assess its value for prognostic risk estimation, and also to validate any findings. Furthermore, development and validation of reliable ways to measure the novel factor might be requested.

Example: Figure 1A in Go *et al* illustrates the association of lower estimated glomerular filtration rate (eg, FR; corresponding to a poor kidney function) with increased risk for death.⁵⁸

3. Prognostic model research (multiple factors of interest)

While a single factor may have limited ability to predict an outcome, several factors together may increase this ability. Prognostic model research (figure 1) focuses on developing a new model or improving or updating an existing model that incorporates several prognostic factors and yet can be easily applied in daily patient care.¹² In addition to evaluating and validating the prognostic ability of models, prognostic model research also aims to assess their clinical utility.

Example: Tangri *et al* developed a model containing four prognostic factors of CKD progression and kidney failure.¹³

4. Stratified medicine research

Since the ultimate goal of prognosis research is to advise patients and guide treatment decisions, the aim of this research is to tailor these decisions using prognostic models in combination with knowledge on factors predicting treatment effects ('predictive factors') and thereby facilitating stratified medicine.¹¹

further testing and are not usefully incorporated into prognostic models.^{4 19–22} Methodological issues that limit the validity and reproducibility of findings often impede the clinical application of prognostic instruments and prevent consensus on its utility that is usually achieved by combining relevant studies in a systematic review.^{5 19 20 23–30}

Various working groups provide guidance and solutions to improve prognosis research. The PROGNosis REsearch Strategy partnership aims to improve all facets of prognosis research (<http://prognosisresearch.com>),^{31 32} and the STRengthening Analytical Thinking for Observational Studies (STRATOS) offers guidance for observational studies (<http://www.stratos-initiative.org/>).³³ STRATOS divides different aspects of observational studies into topic groups (TGs) such as 'TG1: missing data', 'TG5: study design' or 'TG8: survival analysis'.

On behalf of the topic group 'TG5: study design' of the STRATOS initiative, this paper aims to provide guidance for a non-specialist readership on design issues for prognostic factor studies, using four published studies to illustrate key concepts and issues (table 1).^{34–38} For readers less familiar with statistical and other terms, the article is accompanied by a glossary with terms which are italicised in the following text (online supplemental information). In any case, we strongly advise everyone with less statistical expertise to consult researchers with respective expertise (eg, biostatisticians).

In the remainder of this paper, we focus on studies assessing the *prognostic value* of a single factor at a time for a prespecified *outcome*. Moreover, we limit our attention to *binary outcomes* such as the occurrence of KF or death in CKD patients by a particular time-point.

AIMS OF PROGNOSTIC FACTOR STUDIES

Any *prognostic factor* study should be conducted specifically in a sample of the diseased population for which prognosis is intended (so-called target population). While the general aim of *prognostic factor* studies is to identify individual factors that predict patients' future *outcome* (box 1), the aim of a single *prognostic factor* study is far more specific and depends on the stage of research with respect to the factor of interest. Consonni *et al*, for example, aimed to estimate risk for lung cancer recurrence depending on the stage and treatment of the disease (table 1).³⁷ To provide a brief overview:

Discovery studies aim to identify novel factor(s) by screening one or more factors for an association with the *outcome*. Recent studies increasingly use laboratory techniques that permit high throughput measurements from samples from large populations and that allow measurement of many features on each sample (high-dimensional 'Omics' data, eg, genomics data).³⁹ A factor that is associated with the *outcome* (so-called *risk factor*) can be useful for prediction whether or not it is causally related to the *outcome* event occurrence.⁴⁰

However, establishing an association of the factor with the *outcome* is not sufficient in itself to demonstrate its usefulness in informing prognosis (ie, in estimating individual *outcome* risk).^{41 42} In addition, the *prognostic value* of a factor and eventually its *clinical utility* depends on the prevalence/variability of the factor itself as well as on the incidence of the *outcome*.⁴³ For example, a rare *risk factor* may have little use in prognosis. Another aim of the

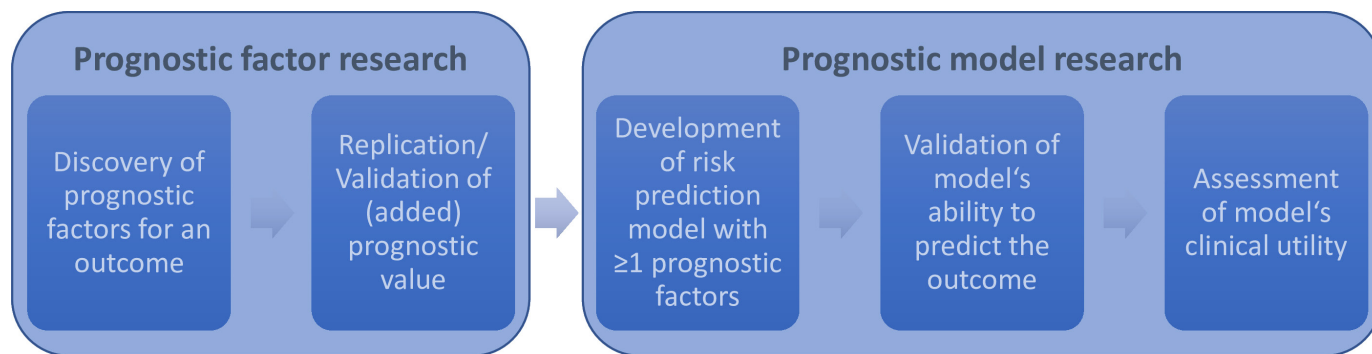


Figure 1 Prognostic factor: illustration of stages in research. See [box 1](#) for more details.

prognostic factor study is thus to assess the *prognostic value* of a factor, especially in addition to established *prognostic factors* (so-called added or incremental *prognostic value*).

Any finding of a *prognostic factor* study should be subjected to internal and external *validation*.^{44 45} In contrast to *internal validation* of results, the *external validation* uses independent data (eg, obtained by a different study group) to check on the validity of observed results. Such attempts may even justify a separate study, in which the *generalisability* of the results could also be evaluated. An example for such a study is reported by Tangri *et al* evaluating the KF risk equation in different geographic regions and different populations and presenting results from combined individual participant data (IPD) in a meta-analysis.^{46 47}

All in all, more than one study is usually required to establish a new *prognostic factor*. Indeed, a systematic review of published studies together with a meta-analysis is a state-of-the-art method for synthesising evidence.³⁰ Although more complex, a meta-analysis using IPD is preferable because it may avoid problems that can arise when combining summary statistics.^{30 47}

Only when a novel factor has proven to be prognostic and can be measured reproducibly, its inclusion into a prognostic model for clinical application may be considered ([box 1](#): prognostic model research).^{12 48} For novel measurements, even additional aspects need to be considered, such as the development of reliable assays to measure them for research purposes, but ultimately also for clinical application.

DATA AND MEASUREMENTS

When designing a *prognostic factor* study, it is critical to identify the data needed to answer the research question. At a minimum, accurate information on the factor and the *outcome* of interest is needed, as well as on previously established *risk* and *prognostic factors* of the *outcome of interest*. Clear definitions and well-defined data collection procedures are required for each data item. Standard operating procedures for data collection and quality control can reduce problems with data such as systematic errors in measurements, and thereby ensure high quality.⁴⁹

Prognostic factor studies require details of the measurement of the factor(s) of interest, which should be done at least once at the beginning of follow-up (baseline). To further improve prediction of *outcome*, longitudinal measurements of certain factors (ie, serial measurements) may be useful, as has been, for example, recommended for carcinoembryonic antigen (CEA) for colorectal cancer.⁵⁰ If one only takes CEA measurements when a patient has abdominal pain following treatment for colorectal cancer, one will overestimate the association of elevated CEA with colorectal cancer recurrence (see chapter on ‘Some Statistical Methods for Immunodiagnostic Cancer Tests’ in Herberman and Mercer).⁵¹ Longitudinal measurements may allow dynamic predictions.⁵²

For some *prognostic factors*, such as serum creatinine, high-quality standardised measurements are available which should be used. However, standardised procedures may not be available for novel factors of interest, which jeopardises the reproducibility of findings from a single study and comparability across studies.^{21 53 54} While the development of a standardised measurement procedure may come later, the basic measurement techniques should be clearly defined *a priori* to promote the internal validity of the measurements.^{10 53 55} In particular, laboratory quality measures (eg, coefficient of variation) and basic description of observed measurements should be provided for any novel factor. A pilot study is useful to demonstrate analytical validity, reproducibility and robustness of measurements.^{53 54 56}

All data should be recorded on the original scale. Categorisation of *continuous* factors should be avoided when collecting data to prevent loss of information.^{25 57} To avoid mistakes, composite measures that are derived from several single measurements (eg, body mass index is derived from height and weight) should be calculated using a programmed algorithm after data collection from the primary data. Measurements that rely on subjective assessments, such as the rating of staining intensity on a slide or the estimation of the percentage of fat tissue on a mammogram, are subject to rater variability. Using more than one rater, each blinded to the source of the biospecimen, can reduce misclassification and bias.⁵³

Table 1 Four published prognostic factor studies

Publication	Albertsen et al (1998; 2005) ^{34, 35}	Variotto et al (2008) ³⁶	Consonni et al (2015) ³⁷	Huzell et al (2015) ³⁸
PMID	9749479; 15870412	18260087	25802059	25556354
Research question regarding prognostic factor(s)	<i>p. 975 Abstract (1998)</i> 'To estimate survival based on a competing risk analysis stratified by age at diagnosis and histologic findings for men diagnosed as having clinically localised prostate cancer...' <i>p. 2095 Abstract (2005)</i> 'To estimate 20-year survival based on a competing risk analysis of men who were diagnosed with clinically localised prostate cancer and treated with... stratified by age at diagnosis and histological findings'.	<i>p. 1547 Abstract</i> '...to assess whether disease-specific survival (DSS) and overall survival (OS) differed among patients who had N1 and N2 bronchioloalveolar carcinoma (BAC) compared with patients who had non-BAC non-small cell lung cancer (NSCLC)'.	<i>p. 1 Abstract</i> '...evaluated prognostic factors for recurrences and survival in 2098 lung cancer case patients ... conducted survival analyses and estimated absolute risks separately for stage IA to IIIA surgically treated and stage IIIB to IV non-surgically treated patients'.	<i>p. 505 Abstract</i> '...was to study oral contraceptive (OC) use in relation to breast cancer events... in a prospective population-based cohort...'
Disease of interest	Prostate cancer, clinically localised	Non-small cell lung cancer (NSCLC), N1/N2; lymph node status	Lung cancer	Primary breast cancer
Exposure of interest	Age at diagnosis Gleason score	Bronchioloalveolar carcinoma (BAC, binary)	Stage IA-IV treatment surgical (ST), non-surgical (NST)	OC (binary)
Study design/study type	Registry-based study, retrospective ▲ Connecticut Tumour Registry (CTR) ▲ Hospital records from 36 acute care hospitals and 2 Veteran Affairs medical centres	Registry-based study, retrospective ▲ source: SEER registry ▲ data access: 3 May 2006 <i>p. 1548 Materials and Methods</i> 'The case ascertainment rate from the SEER registries has been reported to be 97.5% and is believed to be a general representation of the entire American population'.	Add-on study ▲ ongoing case-control study ▲ participants are followed prospectively ▲ Lombardia, Italy <i>p. 2 Methods</i> 'Thus, the enrolled patients were quite representative of all the case patients'.	Add-on study ▲ ongoing prospective cohort study ▲ start in October 2002 ▲ Lund, Sweden
Recruitment period	1971-1984	1992-2002	2002-2005	2002-2011
Inclusion and exclusion criteria	Patients: 1. with prostate cancer 2. aged 55 to 74 years at diagnosis 3. without metastases	Patients: 1. with NSCLC, stages I-IIIb 2. who underwent definitive surgical procedure	Patients: 1. with lung cancer 2. diagnosis Apr 2002-Feb 2005 3. defined catchment area in Lombardia	Patients: 1. with primary breast cancer 2. not preoperatively treated 3. without carcinoma in situ 4. without metastatic spread within 3 months of surgery

Continued

Table 1 Continued

Publication	Albertsen et al (1998; 2005) ^{34, 35}	Varlotto et al (2008) ³⁶	Consonni et al (2015) ³⁷	Huzell et al (2015) ³⁸
Groups of comparison, study size	<ul style="list-style-type: none"> ▶ Gleason score: 2–4: n=138; 5: n=118; 6: n=294; 7: n=137; 8–10: n=80 ▶ Age at diagnosis: 55–59: n=54; 60–64: n=141; 65–69: n=242; 70–74: n=330 <p>In total: 767</p>	Unmatched data set <ul style="list-style-type: none"> ▶ Group: BAC (n=684) (N1: n=408; N2: 276) ▶ Group: non-BAC NSCLC (n=9809) (N1=5626; N2:4183) <p>In total: 10 493 (matched set: 646)</p>	<ul style="list-style-type: none"> ▶ Stage IA–IIIA: 1018 (ST n=760, NST n=258) ▶ Stage IIIB–IV: 1032 (ST n=61, NST n=971) ▶ Stage X: 48 (NST) <p>In total: 2098</p>	<ul style="list-style-type: none"> ▶ Never use (n=276), ever use (n=671) (OC-status missing for one patient) ▶ OC use before age 20 (yes - n=305, no - n=643) ▶ OC use before first child (yes - n=467, no - n=478) <p>In total: 948</p>
Follow-up	1998: follow-up until Mar 1997, mean follow-up: 8.6 years 2005: extended follow-up until Dec 1984, median follow-up (range): 24 years (16–33)	Matched pair set: median 29 months	Range of median follow-up by stage and treatment: 0.7–4.1 years, end of study: Dec 2010	Median follow-up 3.03 years (IQR 1.93–5.23)
Outcome(s) of interest	Death, any cause and cause-specific	OS, DSS	OS, recurrence	breast cancer events (eg, local recurrence, distant metastasis)
Statistical methods	Estimation of probability of dying from prostate cancer or other competing causes, regression-based competing risk analysis	Log-rank test, Kaplan-Meier curves	Kaplan-Meier analysis, log-rank test, Cox regression (HR), Estimation of mortality and recurrence rates, Estimation of absolute risks	Kaplan-Meier analysis, Cox regression (HR, adjusted)
Number of events	Until March 1997: 610 deaths <ul style="list-style-type: none"> ▶ Cause of death known for 553 Mar 1997–Dec 1984: 107 further deaths ▶ In total: 717 deaths ▶ If missing (n=25), cause of death was imputed <p><i>Not reported</i></p>	<i>Not reported</i>	1692 deaths (including 78 from other causes), 633 recurrences (IA–IIIA: 375; IIIB–IV: 258)	100 breast cancer events during follow-up
Sample size or power	<i>Not reported</i>	<i>Not reported</i>	<i>Not reported</i>	<i>Not reported</i>
Validation/replication	p. 957 <i>Introduction (1998)</i> ‘... and (4) a sample size sufficiently large to permit stratification by the biopsy Gleason score and age at diagnosis...’ <i>Not reported</i>	<i>Not reported</i>	<i>Not reported</i>	p. 513 <i>Study limitations</i> ‘The median follow-up time was relatively short’. <i>Not reported</i> p. 513 <i>Study limitations</i> ‘Since this is an exploratory study... the results need to be confirmed in an independent material’. <i>Not reported</i>



THE CHOICE OF OUTCOME

The choice of study endpoint (*outcome*) is critical for answering the research question and other aspects of the study design. Some endpoints are important for defining pathophysiology, while others are chosen for their relevance to clinical management.

As an example, consider the endpoint all-cause mortality. This endpoint is clearly clinically relevant. However, due to the diversity of causes of death, the *prognostic value* of predicting all-cause mortality may not be as useful for clinical care as cause-specific mortality.²⁶ CKD patients, for example, suffer from a higher overall risk of death.³ However, CKD patients often die from causes other than untreated KF, such as cardiovascular events.⁵⁸ Depending on the research question, endpoints such as KF requiring kidney replacement therapy or cardiovascular mortality might be more meaningful.

In the four examples in [table 1](#), the authors considered at least one disease-specific *outcome* such as disease recurrences or disease-specific survival, which are events observed over several years.^{34–38} For more information on the definition of clinical *outcomes*, see also Box 3 in the REMARK guidelines paper.⁵⁹

HOW TO SPECIFY AND SUMMARISE THE OUTCOME APPROPRIATELY?

While the validity and quality control of *outcome* data are just as important as with other data, another important aspect relates to the specification of exactly how the *outcome* is to be measured and analysed. For *outcomes* such as the occurrence of KF or death in CKD patients, data can be understood as a *binary* event which occurred within a defined follow-up period (eg, 1 year). Alternatively, the *time-to-event* can be in the focus of a study.⁶⁰ *Time-to-event* information includes not only whether and when an event (eg, KF) has occurred in a given period after the patient entered the study (time 0=baseline), but also the time of study closure (so-called administrative censoring), or whether and when a specific participant was lost to follow-up. In many settings, other events ('competing risks') also occur in the course of a study that prevent the occurrence of the event of interest.⁶¹ In a study with the *outcome* KF, for example, a competing event would be the death of individuals from non-KF causes such as myocardial infarction.

To summarise *outcome* data, an important prognostic quantity is the probability that the event of interest will occur in a defined time period, such as within 5 years of diagnosis or study inclusion (so-called '*pure risk*'). Often, to account for administrative censoring or loss to follow-up, this probability is estimated from survival curves (eg, using Kaplan-Meier estimator).⁶⁰ Survival curves are functions of time that describe the probability of not having the event of interest at or before a time *t*. Since competing risks reduce the chance that the event of interest will occur, the more clinically relevant type of risk is called *absolute risk*.^{61–63} The differentiation between

absolute and *pure risks* is especially important for long-term prognosis, whereas they are often similar over short time intervals. Special methods are available to compute *absolute risk* (eg, estimating cumulative incidence function).^{61 62}

Instead of using *estimates* of *pure* or *absolute risk*, investigators sometimes use median survival or restricted mean survival to measure prognosis.⁸ However, because survival times (or more generally *times-to-events*) are variable, even for individuals with a similar underlying risk profile, the cumulative probability of the occurrence of the *outcome* to a given time point (ie, *absolute risk*) may be more valuable for guiding clinical management than a statement, for example, about median survival.⁶⁴

WHAT IS THE ESTIMAND OF INTEREST FROM PROGNOSTIC FACTOR STUDIES?

Before the design of a *prognostic factor* study can be determined, the measure of interest (so-called *estimand*) one wants to estimate in the study with a given aim needs to be specified.

For a *prognostic factor* study with the aim of identifying the association between the potential *prognostic factor* and the *outcome* of interest, useful measures of association are *relative odds* (OR), *relative risks* (*risk ratio*) or *relative hazards* (HR).^{62 65} The interest may lie on the estimation of associations in the absence (so-called unadjusted) or in the presence (so-called adjusted) of known *prognostic factors*.

If the aim of the project is to assess the *prognostic value* of a factor by estimating *pure* or *absolute risks*, other measures might be used to evaluate the performance of risk estimations with respect to *calibration* (using *calibration* plots), *discrimination* (measured by the *area under the receiver operating characteristic curve*) or mean squared *prediction error* (called the Brier score for *binary outcomes*).^{66–70} These measures can also be used to assess the added value of a *prognostic factor* by comparing corresponding measures derived from a model that includes only known *prognostic factors* with a model that additionally includes the factor of interest.^{66 68 70} Further, measures such as *R*² can also be used to assess model's goodness-of-fit to describe the *outcome*.^{66 70}

STUDY DESIGNS FOR PROGNOSTIC FACTOR STUDIES

After clarification of the study aim and the *estimand* one seeks to estimate, an appropriate design can be chosen taking into account the medical context of the disease (eg, treatments).

Because factors of interest are usually not subject to experimental control as they are either non-modifiable (eg, genetic predisposition) or *risk factors* with potential detrimental health effects and thus unethical to administer to participants of a study (eg, smoking), *prognostic factor* research is mainly done on observational data.⁶⁵

If the purpose of the study is simply to determine whether a factor is associated with an *outcome*, a *case-control*

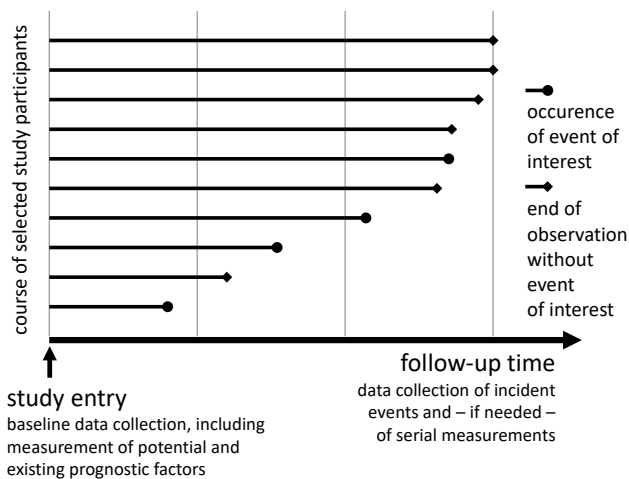


Figure 2 Conceptual illustration of a cohort study for use in prognostic factor research. In the case of a *prospective cohort study*, the study entry lies in the present, whereas in a *retrospective cohort study*, the cohort is assembled retrospectively from information collected in the past. Of note, the calendar date of study entry of the participants does not necessarily have to be the same. Nevertheless, participants in both types of studies are followed as time goes on starting at the time of inclusion. For retrospective cohort studies, follow-up data may still be available from the past, but data collection may also continue in the future. An event of interest may be any prespecified outcome such as kidney failure in patients with chronic kidney disease. End of observation may include censoring due to administrative closure of the study or the loss of contact with some participants (lost to follow-up) or the occurrence of a competing event such as death from causes other than kidney failure.

study design may be appropriate.⁶⁵ Comparison of factor's distributions between individuals selected from the target population by *outcome* status (presence, cases; absence, controls) leads to estimates of the *relative odds* of disease per unit increase in the *risk factor* (if *continuous*) or versus the reference category (if *binary* or *categorical*). Its calculation from case-control data theoretically corresponds to the *relative odds* of the *outcome* that can alternatively be obtained from a *cohort study*.⁷¹ *Case-control studies* have limitations.^{65 71} Major issues of case-control designs in *prognostic factor* research include ensuring that the prognostic variables were reliably measured and correspond to the start point at which prognostic information was required. Accounting for censored observations is also problematic.

Alternatively, but also if the purpose of the study is to estimate *absolute* or *pure risk*, the gold-standard design is a *prospective cohort study* (figure 2).^{62 65} For this purpose, participants are recruited from the target population for whom data are available on factor(s) that were measured at or before the starting point of interest for prognostication. Although restrictive study inclusion and exclusion criteria may yield a homogeneous study cohort useful for identifying new *prognostic factors*, a more representative cohort can yield risk predictions that are more accurate for the target population.⁷² Of

advantage, *outcomes* can be completely and accurately assessed during the follow-up of this cohort. Important aspects to consider when designing a cohort study are the timing of visits and the length of follow-up. The follow-up should be conducted on a fixed schedule for all study participants, regardless of their clinical status, to avoid selection bias or informative missing data, and to mirror clinical practice.^{53 73} In the end, a *cohort study* yields information on measures of association such as *relative odds* or *relative hazards* as well as on *pure* and *absolute risks*.

Overall, an advantage of a *prospective cohort study* tailored to the purpose of *prognostic factor* research (ie, data collected specifically for the condition and *outcomes* of interest) is that one can collect high-quality data including previously established *prognostic factors* and treatment. However, a *prospective cohort study* is often costly and takes a long time to observe a sufficient number of events of interest during the follow-up.^{10 53 65}

Using already available data and biospecimens: pros and cons

To avoid costs and time issues of *de novo prospective cohort studies*, researchers have taken advantage of already collected data and biospecimens in the past.^{7 24 74}

Prognostic factor research can also be conducted as an add-on project to an existing *prospective cohort study* whose study population resembles the target population of interest.^{53 75} The existing study may have standardised procedures for collection of baseline data and biospecimens and for *outcome* ascertainment, and it may be set up to support add-on studies, saving time and expense. The German Chronic Kidney Disease study is an example of an ongoing prospective observational study of patients with moderately reduced kidney function at baseline providing a supportive structure for add-on projects.⁷⁶

If required data are available on all members of the cohort, standard methods can be used.^{8 60 62} However, required data are sometimes only available for a subcohort when, for example, it is too expensive to analyse biospecimens from all members of the cohort to obtain measurements on the factor of interest. Fortunately, one can still estimate *absolute* or *pure risk* by obtaining such factor data from cases and a subsample of non-cases using a *nested case-control* or *case-cohort* design.^{62 65} These subsampling strategies can greatly reduce the cost of prognosis research but still require other baseline data from all cohort members.

Randomised clinical trials include prospective cohorts that are also potentially useful for observational prognosis research. Their specific advantages are that the initial treatment regimens are known, the population is well defined and the baseline data and follow-up data are usually of high quality.⁷⁵ However, *estimates* of *pure* or *absolute risk* from clinical trial populations may not be *generalisable* to the target population of interest if the clinical trial population is highly restricted by eligibility criteria or if the follow-up procedures for ascertaining events differ from those in common use in the target population.



Other secondary data sources, like registries, databases, electronic health records (EHRs) or biospecimen repositories may also provide a basis to retrospectively assemble a cohort (often called: *retrospective cohort study*; figure 2) for prognosis research.⁵³ Some sources are dedicated to specific diseases or groups of diseases, such as the German Centre for Cancer Registry Data, which obtains data from all cancer patients in Germany (https://www.krebsdaten.de/Krebs/EN/Home/homepage_node.html). Other sources like EHRs usually cover a larger population but may not be large enough to include enough people with the disease of interest.

While secondary data sources are very convenient, they may have disadvantages—especially in the case of sources like registries, databases or EHRs.^{5 10 19 24 53} Among others, the source population for people with the disease of interest of a registry, database or EHRs can be less well-defined than in studies with specific entry criteria and thus limit its representativeness for the population that is targeted in the *prognostic factor* study. Ideally, the representativeness of any secondary data source is assessed before using it for prognosis research. In addition, the extent and definitions, as well as the timing and quality of data covered in these sources, might be limited, less complete and not rectifiable.^{7 72}

The four studies highlighted in table 1 are examples of studies leveraging already collected data. Varlotto *et al* used data from an established cancer registry in the USA (Surveillance, Epidemiology, and End Results Programme database, <https://seer.cancer.gov/>), for which a case ascertainment rate of 97.5% in detection areas was reported.³⁶ The authors thus provided evidence for the representativeness of underlying data.

Albertsen *et al*^{34 35} also present a registry-based study (table 1) that identified patients with prostate cancer between 1971 and 1984 in the Connecticut Tumor Registry. Registered patients had been followed according to registry protocols to determine the date and cause of death.³⁴ Hospital records of eligible patients were additionally reviewed to successfully online supplemental information available in the registry.

In summary, there are clear advantages to add-on studies by leveraging previously gathered baseline data, biospecimens and follow-up information. However, these advantages may not outweigh limitations of these data sources. Thus, before deciding to base prognosis research on an existing study or registry, one must evaluate the appropriateness of the data source and consider its limitations. Disadvantages may be acceptable in early stages of *prognostic factor* research (eg, *discovery*; figure 1), but a *prospective cohort study* tailored to the target population is preferred in advanced *stages of research* (box 1).

STATISTICAL METHODS

Depending on study aim, *estimand* of interest and design, the statistical analysis should be planned accordingly.

In *discovery* studies with a case-control design, for example, standard *statistical tests* might be used respectively to compare the distribution of a *continuous* factor or *binary* factor in cases and controls. However, fitting *regression models* to case-control but also to cohort data is preferred to account for other known *prognostic factors* which traditionally are known as confounders in observational studies, but here are merely other *prognostic factors*.^{8 71 77} While the estimation of unadjusted associations is valuable to describe the raw dependency of the factor with the *outcome*, adjustment for known prognostic factors may indicate whether the factor carries additional independent information about the *outcome*. This adjustment for existing *prognostic factors* allows the researcher to examine the bespoke factor's added prognostic value, over and above that from the existing *prognostic factors*. Various *regression models* are available. For example:

A logistic *regression model*, for example, can be used to describe the (adjusted) association of a factor in terms of the *relative odds* of the *outcome* with and without the factor in a *case-control study*.⁷¹ When censoring, loss to follow-up and competing events are negligible in a cohort study, as in cohort studies with short follow-up times (eg, a study of death from heart disease within 1 month of a myocardial infarction), a logistic *regression model* can also be used to estimate (adjusted) ORs.^{63 71} For *time-to-event* data of a *cohort study*, *survival methods* (eg, Cox proportional hazard *regression model*) can be used to estimate (adjusted) associations (*relative hazards*).⁶⁰

Regression models allow for the derivation of *absolute risks* of individuals, necessary when aiming to assess the *prognostic value* of a factor.^{62 63} A necessary ingredient is the baseline risk of the *outcome*. In contrast to cohort studies, *case-control studies* do not allow their direct estimation, so that the information must be obtained from published data.⁷⁸ However, when transferring back to the *absolute risk* scale, care should be taken to account for the competing event to reflect a real world situation where the competing event may occur.^{61 63}

Furthermore, assumptions underlying every *regression model* (eg, Cox model: proportional hazard assumption⁶⁰) need to be checked. Usual assumptions of linearity and additivity refer to how variables are selected for inclusion in the model.⁷⁹ This includes decisions on how *categorical* variables with three or more levels are coded and whether a *continuous* variable can be included on its original scale.⁸⁰ A *continuous* variable should not be categorised, unless for biological reasons.^{25 57 63}

Published *prognostic factor* studies report a mixture of (adjusted) effect sizes in terms of *odds*, *risk* or *hazard ratios*, and *pure/absolute risks* at particular time points. Consider the four examples of *prognostic factor* studies in table 1. Huzell *et al*, for example, assessed the effect of oral contraceptive use (eg, use before age 20 years—no/yes) on the prognosis of breast cancer patients and reported, among others, measures of association (*hazard ratios*) with breast cancer events (eg, recurrence) using a Cox proportional hazard *regression model*.³⁸ In comparison,

Albertsen *et al* estimated survival in male patients with prostate cancer depending on age at diagnosis and histological findings.^{34 35} By differentiation of the *outcome* of interest (death from prostate cancer) from death from other causes (competing risk), the authors provide *estimates* of cumulative mortality (*absolute risks*) across various ages and histological findings as illustrated in figure 1 of Albertsen *et al*.³⁴

In *discovery* studies with high-dimensional omics data, several approaches are available. Researchers may wish to evaluate each omics feature separately. In this situation, there is not much difference in the procedure compared with a study that evaluates only one factor—apart from the burden of evaluating so many features in parallel. To avoid too many falsely positive results (ie, inflated type I error), methods to adapt the significance threshold have become standard.^{81–83} Another approach to deal with many features is to reduce dimensionality by either selecting a subset of features or by defining substitutes supposedly carrying the relevant information of whole omics data. Regularised/penalised regression methods (eg, Lasso regression) and machine learning/artificial intelligence approaches (eg, random forest, neural networks) are methods of choice.^{79 84 85}

More detailed literature on statistical methods, their application and recommendations can be found elsewhere.^{8 32 59 62 63 71 86} Of note, a special procedure is required when aiming to validate the *prognostic value* of a factor or model.⁸⁷

SAMPLE SIZE AND POWER OF A PROGNOSTIC FACTOR STUDY

Assuring adequate power is one of the most important features for the success of a *prognostic factor* study.^{2 7 53} For a *prospective cohort study*, the investigator may be able to estimate sample size to ensure adequate statistical power, whereas the sample size is usually fixed in retrospectively assembled studies. The power of such a study should be assessed instead. Because the number of events reflects the effective sample size driving the power of *statistical tests* to compare proportions and of *time-to-event* analyses, considerations of sample size and power are particularly relevant for less common *outcomes*.^{6–8}

Standard methods to calculate the sample size or evaluate power are available for some aims of *prognostic factor* research.^{88–92} For a *binary* factor and an unadjusted effect of interest, for example, the calculations are similar to those for a clinical trial. Requested inputs are the *relative risk* between groups with and without factor and the prevalence of the factor.⁹⁰ However, where an adjusted effect is required, the sample size needs to account for the correlation with the other factors, which usually inflates the sample size needed.

Since prior knowledge is always required for the calculation of sample size or power, a search of published literature is needed. The result of the calculations can then indicate whether the study is feasible and worth pursuing. If the sample size or statistical power is inadequate to

meet the aims of the study, the study should not proceed without revision of the design, because otherwise the study is likely to produce misleading or useless results and a waste of biospecimens and investigator resources. Waste of valuable biospecimens raises ethical concerns.^{23 88 93}

Researchers could consider increasing the sample size by collaborating with researchers at other institutions. Collaborative projects, however, require additional planning for data harmonisation and statistical analysis. Results from such endeavours could be reported from the analysis of combined IPD (preferred) or meta-analysis combining summary statistics from contributing institutions.^{30 47} Such a collaboration would also have the advantage of (considerably) reducing problems of heterogeneity which frequently arise due to various differences between studies and their analyses (eg, variables used for adjustment).^{30 94}

In practice, sample size considerations are rarely reported in *prognostic factor* research, and many such studies, especially studies leveraging existing data, are too small.^{87 95 96} The four publications in table 1 did not report sample size or power calculations.^{34–38} Some would argue that after the study is completed, quantities such as confidence intervals for *relative hazards* or *absolute risks* indicate whether the sample size was sufficient. However, a significant result might be a false positive, making judgements from such quantities vulnerable.^{97 98} In addition, low power also results in more incorrect non-rejection of null hypotheses (false negative). Therefore, considerations regarding the power of a study should be made before starting the study and then also reported.⁹⁰

STUDY PROTOCOL AND REPORTING

A carefully written protocol is an essential element of a good design of a *prognostic factor* study.^{2 5 26 99} Writing the protocol by the study team helps assure that the aims, definitions and methods are understood and agreed to.

Templates for different study types are available, such as the harmonised protocol template of the International Society for Pharmacoepidemiology and the International Society for Pharmacoeconomics and Outcomes Research.³⁹ They share basic elements. In addition to aspects discussed in this article, background information justifying the necessity of the study and its aims is also required, which can be obtained through a literature review. For the convenience of readers, we assembled aspects and questions to address when designing a *prognostic factor* study in table 2 that may even serve as a guideline or check list. In addition, Figure 3 illustrates the complexity of the process when designing a *prognostic factor* study, which is due to the dependencies of individual aspects. Failures in design or measurements are often irreparable by statistical analysis.

A protocol not only serves as a guide to the conduct of the study, but it is also useful for communicating the importance and quality of the study to potential funders and as a point of reference when writing publications

**Table 2** List of general aspects to consider when designing a study for prognostic factor research**A clear definition of the research goals and objectives is informed by:**

Task	Aspects to consider
Literature search	<ul style="list-style-type: none"> ▶ What information is already available? ▶ What is the status/stage of research? ▶ What is known about the factors to be studied (eg, distribution, variability)? ▶ Which outcome(s) are clinically relevant? ▶ What are the prevalences or incidences of the index disease and outcomes? ▶ Are there other factors that influence outcomes (eg, treatment, known prognostic factors)? ▶ What are major limitations of past studies? ▶ Are there data sources that might be used?
Estimand	<ul style="list-style-type: none"> ▶ What is the measure of interest? ▶ With respect to stage of research and aim of intended study, discuss the use of relative measures (measures of association) versus absolute measures allowing to evaluate prognostic value of a factor or model.
Definition of population and major variables	<ul style="list-style-type: none"> ▶ What is the target population of interest? ▶ Provide unambiguous definitions of the index disease, exposure(s) and outcome(s). ▶ Compare definitions applied in past studies for consistency with definitions in the planned study. ▶ Define the scale on which variables should be collected. ▶ Specify variables that need to be derived from collected data.
Selection of the appropriate study design	<ul style="list-style-type: none"> ▶ Discuss advantages and limitations of available data sources to choose one that best meets the aims and design of the new study (eg, resemblance to target population, definition, quality and completeness of data on key variables, sample size). ▶ Define a study design appropriate for the stage of research. ▶ Choose inclusion and exclusion criteria that meet study aims without threatening generalisability.
Study size/power	<ul style="list-style-type: none"> ▶ Define sample size and length of follow-up to yield the number of events needed for adequate power. ▶ Account for loss of patients during follow-up in estimating required number of events. ▶ If the available sample size is fixed, discuss power to see if the study is feasible. ▶ If the sample size and power are not sufficient, consider a multicentre approach.
Data and data collection	<ul style="list-style-type: none"> ▶ Define data collection procedures to promote accurate, reproducible, complete data, especially for the key variables. ▶ Develop procedures to detect errors in the data and means to correct them. ▶ Consider collecting data on other variables of interest, such as treatment, known confounding factors and known prognostic factors. ▶ For novel measurements, ascertain quality of measurements when no report exists.
Study protocol	<ul style="list-style-type: none"> ▶ Compile a carefully written protocol describing rationale, aims and the items above. ▶ Describe measures of quality control and initial data analysis of collected data. ▶ Present statistical analyses tailored to study objective. Include sensitivity analyses, as needed. ▶ <i>Recommended:</i> registration of the study protocol.
Additional quality control checks	<ul style="list-style-type: none"> ▶ Is prior knowledge sufficient to justify the study aims and design? ▶ Are data on key variables accurate and reproducible, and is there data on the validity of measurements on novel variables to be studied as prognostic factors? ▶ Is there a plan to train study personnel to ensure quality and completeness of study data? ▶ Are serial samples for prognostic factor measurements obtained independently of the health status of the participant? ▶ How will completeness of data collection and follow-up be promoted and assessed? ▶ Discuss options for external validation of findings.

Listed aspects need to be considered together as they are not independent of each other and may have influence on subsequent decisions.

If any aspect cannot be addressed sufficiently, a pilot study or simulation study might be useful to detect problems and missing information, and thereby overcome issues before launching the main study.

Moreover, as a *general recommendation*: Consult an experienced biostatistician when designing a study!

based on the study. Due to poor reporting of study results in the past,^{100 101} reporting guidelines for different types of studies have been developed and can be found at the homepage of Equator Network (<https://www.equator-network.org/>). The reporting recommendations for

tumour marker prognostic studies (REMARK) are highly pertinent to *prognostic factor* research.¹⁰² Another pertinent reporting guideline is the TRIPOD-AI guideline.¹⁰³ Although its focus is more on prognostic model research than on *prognostic factor* research, many of the listed aspects

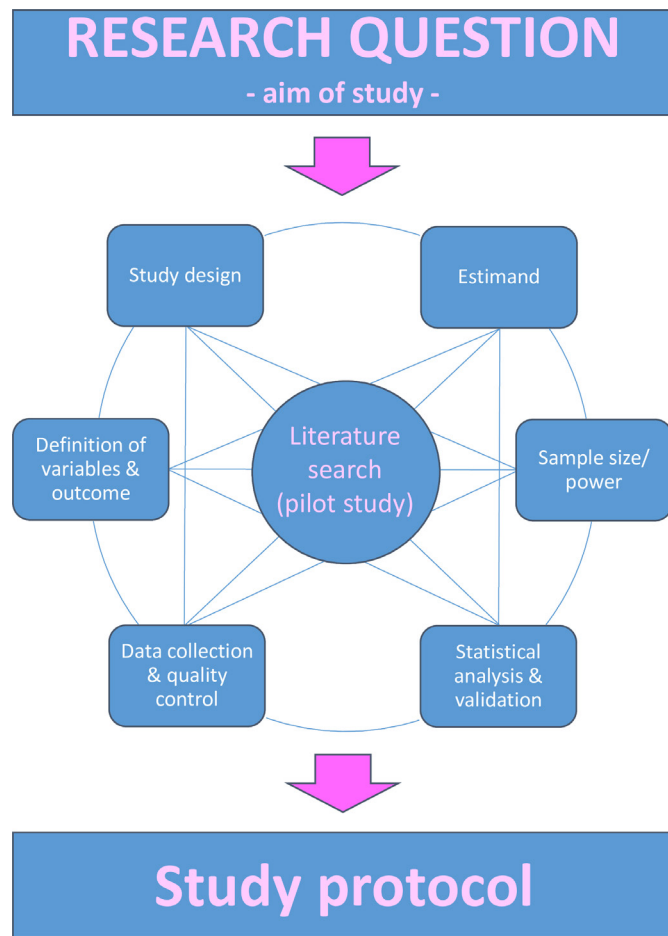


Figure 3 Conceptual illustration of the inter-related aspects in designing a prognostic factor study. The figure attempts to illustrate aspects involved and their dependencies when designing a prognostic factor study. See [table 2](#) for more details on single aspects.

are relevant to both and even provide important information on the *validation* of results. Detailed explanations and examples have been provided for several reporting guidelines in separate publications (‘Explanation and Elaboration’), including for REMARK and TRIPOD.^{59 104} Although the primary purpose of reporting guidelines is to educate about good reporting, these papers also provide a valuable reference for the many aspects that need to be considered when designing, conducting and analysing studies.

To improve design, methods and reporting in *prognostic factor* research and also to facilitate systematic review on *prognostic factors*, registration in a public database has been advocated.^{5 10 26 99 105–107} Currently, registration is mandatory for clinical trials and post-authorisation safety studies with marketed drugs, but only recommended for other observational studies.

SUMMARY

We have introduced the general researcher to elements of *prognostic factor* research, focusing on the evaluation of a single factor at a time with a *binary* (including *time-to-event*)

outcome to illustrate design and methodological principles. While there are numerous other settings and aspects to consider when designing a *prognostic factor* study, this manuscript cannot address them all. For instance, we did not consider *continuous outcome* measures such as the decline in eg, FR, a state-of-the-art *outcome* measure of kidney function.¹⁴ Also, while aspects were discussed in the context of prospective cohorts with some extension to retrospectively assembled cohorts, other designs (eg, *nested case-control studies*) and data sources (eg, EHR data) may require additional considerations of other aspects. To cover some of the many other and related aspects of prognosis research, it is recommended that the reader consult literature on the development, *validation* and use of prognostic models.^{8 32 62 63} The *external validation* of any result such as the performance of a *prognostic factor* or model and the systematic review of literature is of utmost importance to eventually reach a final conclusion on the *prognostic value* of a factor or model.^{30 87}

Overall, shortcomings in design and methodology lead to wasted efforts and resources and partly explain why only a small proportion of new *prognostic factors* is translated into clinically useful information, eg, as part of prognostic models.^{22 27 108} More time should be invested in designing a *prognostic factor* study to avoid problems that cannot otherwise be resolved in the statistical analysis. Improvements will, however, require the combined efforts of researchers and other stakeholders.^{22 108–110} Although scientific intuition may lead to promising *prognostic factors*, attention to methodology will be a key element for successful translation of results. Indeed, van Calster *et al* argue ‘that the core problem is a paradox: methodology, the very backbone of science, remains overly trivialised by the scientific community that funds, undertakes and reports (pre) clinical research’.¹⁰⁶

Finally, it must also be acknowledged that there is not a complete solution or recommendation for all methodological aspects in *prognostic factor* research (eg, for prioritising factors in an *external validation*). With the emergence of new data sources (eg, EHR) and methods (eg, measurement techniques), experts in the field are asked to continue to strive for solutions and recommendations for old and new problems.

We hope that this introduction to *prognostic factor* research will raise awareness of the need for high-quality studies and of potential methodological issues that could be addressed at the design stage and thus avoided. For support, information included in display items will provide some additional guidance. We encourage researchers planning prognosis research to consult an experienced biostatistician early and regularly.

Author affiliations

¹Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, Germany



²Department of Medicine IV - Nephrology and Primary Care, Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, Germany

³Synlab MVZ Humangenetik Freiburg GmbH, Freiburg, Germany

⁴Real World Evidence, TFS HealthScience, Barcelona, Spain

⁵Bicocca Bioinformatics, Biostatistics and Bioimaging Centre - B4, School of Medicine and Surgery, University of Milan-Bicocca, Milano, Italy

⁶Unit of Clinical Epidemiology and Biostatistics, IRCCS San Gerardo dei Tintori Foundation, Monza, Lombardia, Italy

⁷Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, USA

⁸Leslie Dan Faculty of Pharmacy & Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

⁹School of Health Sciences, College of Medicine and Health, University of Birmingham, Birmingham, UK

¹⁰National Institute for Health and Care Research (NIHR), Birmingham Biomedical Research Centre, Birmingham, UK

¹¹Centre for Statistics in Medicine, UK EQUATOR Centre, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, UK

¹²Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, Germany

¹³Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA

Collaborators Full list of the members of the topic group 'TG5: study design' of the STRATOS initiative: Co-chairs: Mitchell H Gail, Suzanne M Cadarette. Members: Gary S Collins, Susan Halabi, Rima Izem, Thomas Lumley, Paola Rebora, Peggy Sekula, Neus Valveny. Early Career Adjunct Member: Nicholas Bakewell.

Contributors All authors substantially contributed to the conception/design of the work, to drafting and critically revising the content and provided the final approval of the version to be published. The authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Guarantor: PS (corresponding author).

Funding We acknowledge support by the Open Access Publication Fund of the Albert-Ludwigs-Universität Freiburg, Germany. The work of PS was supported by the German Research Foundation (DFG) Project-ID 431984000 - SFB 1453 and SE 2407/3-1. PS received travel grants to attend the 3rd Human Glycom Project meeting (Nov 2022) and the 16th Meeting of the German Society of Nephrology (Sept 2024). The work of IS was supported by the German Research Foundation (DFG) Project-ID 530592017 (SCHL 2292/3-1). The work of UTS was supported (incl. travel grants) by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grant 01ZX1912B). The work of PR was supported by the European Union - Next Generation EU - NRRP M6C2 - Investment 2.1 Enhancement and strengthening of biomedical research in the NHS project code PNRR-MAD-2022-12376033. SH is supported by National Institutes of Health (NIH) Grants R01CA256157, R01CA249279, the United States Army Medical Research Award HT9425-23-1-0393, the Food and Drug Administration (FDA) Award 1U01FD007857-01 of the US Department of Health and Human Services (HHS), and the Prostate Cancer Foundation. SMC is supported by NIH Grant R01AG078759 funded by the National Institute on Aging (NIA), the NIH Office of the Director, and the Office of Disease Prevention (ODP). RDR is supported by the National Institute for Health and Care Research (NIHR) Birmingham Biomedical Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham, and by an MRC-NIHR grant (reference: MR/V038168/1). RDR and GSC are NIHR senior investigators. GSC was supported by Cancer Research UK (programme grant: C49297/A27294). The work of WS was supported by the German Research Foundation (DFG), grant number SA580/10-3. All other authors have nothing to disclose.

Competing interests UTS is working for SYNLAB MVZ Humangenetik Freiburg GmbH. NV is employee of TFS HealthScience. SH serves on several data monitoring committees (CG Oncology, BMS, Janssen, Beigene, Sanofi). RDR is affiliated with an institute that is funded by a grant by NIHR. The funder did not influence the results/outcomes of the study despite author affiliations with the funder. RDR also receives royalties for textbooks on Prognosis Research in Healthcare & Individual Participant Data Meta-Analysis and consultant fees (BMJ Statistics Editor). All other authors have nothing to disclose.

Ethics approval Not applicable.

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

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iDs

Peggy Sekula <http://orcid.org/0000-0003-2263-447X>

Inga Steinbrenner <http://orcid.org/0000-0003-0851-5773>

Susan Halabi <http://orcid.org/0000-0003-4135-2777>

Richard D Riley <http://orcid.org/0000-0001-8699-0735>

Gary S Collins <http://orcid.org/0000-0002-2772-2316>

Mitchell H Gail <http://orcid.org/0000-0002-3919-3263>

REFERENCES

- 1 Thomas JM, Cooney LM, Fried TR. Prognosis Reconsidered in Light of Ancient Insights-From Hippocrates to Modern Medicine. *JAMA Intern Med* 2019;179:820-3.
- 2 Riley RD, Hayden JA, Steyerberg EW, *et al*. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLoS Med* 2013;10:e1001380.
- 3 Jha V, Garcia-Garcia G, Iseki K, *et al*. Chronic kidney disease: global dimension and perspectives. *Lancet* 2013;382:260-72.
- 4 Provenzano M, Rotundo S, Chiodini P, *et al*. Contribution of Predictive and Prognostic Biomarkers to Clinical Research on Chronic Kidney Disease. *IJMS* 2020;21:5846.
- 5 Riley RD, Sauerbrei W, Altman DG. Prognostic markers in cancer: the evolution of evidence from single studies to meta-analysis, and beyond. *Br J Cancer* 2009;100:1219-29.
- 6 Benedetti JK, Liu PY, Sather HN, *et al*. Effective sample size for tests of censored survival data. *Biometrika* 1982;69:343-9.
- 7 Altman DG. Prognostic models: a methodological framework and review of models for breast cancer. *Cancer Invest* 2009;27:235-43.
- 8 Harrell FE. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. 2nd edn. Springer, 2015.
- 9 Gray LJ, Bath PMW, Collier T, *et al*. Should stroke trials adjust functional outcome for baseline prognostic factors? *Stroke* 2009;40:888-94.
- 10 McShane LM, Hayes DF. Publication of Tumor Marker Research Results: The Necessity for Complete and Transparent Reporting. *JCO* 2012;30:4223-32.
- 11 Hingorani AD, Windt DA van der, Riley RD, *et al*. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ* 2013;346:e5793.
- 12 Steyerberg EW, Moons KGM, van der Windt DA, *et al*. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med* 2013;10:e1001381.
- 13 Tangri N, Stevens LA, Griffith J, *et al*. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA* 2011;305:1553-9.
- 14 Pena MJ, Stenvinkel P, Kretzler M, *et al*. Strategies to improve monitoring disease progression, assessing cardiovascular risk, and defining prognostic biomarkers in chronic kidney disease. *Kidney Int Suppl (2011)* 2017;7:107-13.
- 15 Sopyllo K, Erickson AM, Mirtti T. Grading Evolution and Contemporary Prognostic Biomarkers of Clinically Significant Prostate Cancer. *Cancers (Basel)* 2021;13:628.
- 16 Lee AHS, Ellis IO. The Nottingham prognostic index for invasive carcinoma of the breast. *Pathol Oncol Res* 2008;14:113-5.
- 17 Collaborators MCT, Perel P, Arango M. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *BMJ* 2008;336:425-9.

- 18 Califf RM. Biomarker definitions and their applications. *Exp Biol Med (Maywood)* 2018;243:213–21.
- 19 McShane LM, Altman DG, Sauerbrei W. Identification of clinically useful cancer prognostic factors: what are we missing? *J Natl Cancer Inst* 2005;97:1023–5.
- 20 Kern SE. Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. *Cancer Res* 2012;72:6097–101.
- 21 Plebani M. Lessons from controversy: biomarkers evaluation. *Clin Chem Lab Med* 2013;51:247–8.
- 22 Macleod MR, Michie S, Roberts I, et al. Biomedical research: increasing value, reducing waste. *Lancet* 2014;383:101–4.
- 23 Altman DG. The scandal of poor medical research. *BMJ* 1994;308:283–4.
- 24 Altman DG, Riley RD. Primer: an evidence-based approach to prognostic markers. *Nat Rev Clin Oncol* 2005;2:466–72.
- 25 Sauerbrei W. Prognostic factors. Confusion caused by bad quality design, analysis and reporting of many studies. *Adv Otorhinolaryngol* 2005;62:184–200.
- 26 Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. *BMJ* 2009;339.
- 27 Ioannidis JPA, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet* 2014;383:166–75.
- 28 Martino D, Chai-Adisaksopha C, Iorio A. Systematic reviews of prognosis studies: a critical appraisal of five core clinical journals. *Diagn Progn Res* 2017;1:9.
- 29 Sauerbrei W, Haeussler T. Comment on “BAG-1 as a biomarker in early breast cancer prognosis: a systematic review with meta-analyses”. *Br J Cancer* 2018;118:1152–3.
- 30 Riley RD, Moons KGM, Snell KIE, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ* 2019;364:k4597.
- 31 Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595.
- 32 Riley RD, van der Windt D, Croft P, et al. *Prognosis research in healthcare: concepts, methods and impact*. 2019.
- 33 Sauerbrei W, Abrahamowicz M, Altman DG, et al. Strengthening analytical thinking for observational studies: the STRATOS initiative. *Stat Med* 2014;33:5413–32.
- 34 Albertsen PC, Hanley JA, Gleason DF, et al. Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer. *JAMA* 1998;280:975–80.
- 35 Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA* 2005;293:2095–101.
- 36 Varlotto JM, Flickinger JC, Recht A, et al. A comparison of survival and disease-specific survival in surgically resected, lymph node-positive bronchioloalveolar carcinoma versus nonsmall cell lung cancer: implications for adjuvant therapy. *Cancer* 2008;112:1547–54.
- 37 Consonni D, Pierobon M, Gail MH, et al. Lung cancer prognosis before and after recurrence in a population-based setting. *J Natl Cancer Inst* 2015;107:djv059.
- 38 Huzell L, Persson M, Simonsson M, et al. History of oral contraceptive use in breast cancer patients: impact on prognosis and endocrine treatment response. *Breast Cancer Res Treat* 2015;149:505–15.
- 39 Wang SV, Pottegård A, Crown W, et al. HARmonized Protocol Template to Enhance Reproducibility of Hypothesis Evaluating Real-World Evidence Studies on Treatment Effects: A Good Practices Report of a Joint ISPE/ISPOR Task Force. *Value Health* 2022;25:1663–72.
- 40 Tripepi G, Jager KJ, Dekker FW, et al. Testing for causality and prognosis: etiological and prognostic models. *Kidney Int* 2008;74:1512–5.
- 41 Pepe MS, Janes H, Longton G, et al. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 2004;159:882–90.
- 42 Feng Z. Classification versus association models: Should the same methods apply? *Scand J Clin Lab Invest* 2010;70:53–8.
- 43 Foroutan F, Iorio A, Thabane L, et al. Calculation of absolute risk for important outcomes in patients with and without a prognostic factor of interest. *J Clin Epidemiol* 2020;117:46–51.
- 44 Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130:515–24.
- 45 Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013;13:33.
- 46 Tangri N, Grams ME, Levey AS, et al. Multinational Assessment of Accuracy of Equations for Predicting Risk of Kidney Failure: A Meta-analysis. *JAMA* 2016;315:164–74.
- 47 Riley RD, Tierney JF, Stewart LA, eds. *Individual participant data meta-analysis: a handbook for healthcare research*. Wiley, 2021.
- 48 Sachs MC, Sjölander A, Gabriel EE. Aim for Clinical Utility, Not Just Predictive Accuracy. *Epidemiology* 2020;31:359–64.
- 49 Schmidt CO, Struckmann S, Enzenbach C, et al. Facilitating harmonized data quality assessments. A data quality framework for observational health research data collections with software implementations in R. *BMC Med Res Methodol* 2021;21:63.
- 50 Nicholson BD, Shinkins B, Pathiraja I, et al. Blood CEA levels for detecting recurrent colorectal cancer. *Cochrane Database Syst Rev* 2015;2015:CD011134.
- 51 Herberman RB, Mercer DW. *Immunodiagnosis of cancer*. 2nd edn. M. Dekker, 1990.
- 52 Halabi S, Li C, Luo S. Developing and Validating Risk Assessment Models of Clinical Outcomes in Modern Oncology. *JCO Precis Oncol* 2019;3:1–12.
- 53 Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 2009;101:1446–52.
- 54 Rifai N, Watson ID, Miller WG. Commercial immunoassays in biomarkers studies: researchers beware! *Clin Chem* 2012;58:1387–8.
- 55 Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 2001;93:1054–61.
- 56 Anderson DC, Kodukula K. Biomarkers in pharmacology and drug discovery. *Biochem Pharmacol* 2014;87:172–88.
- 57 Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006;25:127–41.
- 58 Go AS, Chertow GM, Fan D, et al. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* 2004;351:1296–305.
- 59 Altman DG, McShane LM, Sauerbrei W, et al. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): Explanation and Elaboration. *PLoS Med* 2012;9:e1001216.
- 60 Perera M, Dwivedi AK. Statistical issues and methods in designing and analyzing survival studies. *Cancer Rep* 2020;3:e1176.
- 61 Austin PC, Ibrahim M, Putter H. Accounting for Competing Risks in Clinical Research. *JAMA* 2024;331:2125.
- 62 Pfeiffer RM, Gail MH. *Absolute risk: methods and applications in clinical management and public health*. CRC Press Taylor & Francis, 2018.
- 63 Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. Springer, 2009.
- 64 Henderson R, Keiding N. Individual survival time prediction using statistical models. *J Med Ethics* 2005;31:703–6.
- 65 Gail MH, Altman DG, Cadarette SM, et al. Design choices for observational studies of the effect of exposure on disease incidence. *BMJ Open* 2019;9:e031031.
- 66 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiol Camb Mass* 2010;21:128–38.
- 67 Schoop R, Beyersmann J, Schumacher M, et al. Quantifying the predictive accuracy of time-to-event models in the presence of competing risks. *Biom J* 2011;53:88–112.
- 68 Parikh CR, Thiessen-Philbrook H. Key Concepts and Limitations of Statistical Methods for Evaluating Biomarkers of Kidney Disease. *J Am Soc Nephrol* 2014;25:1621–9.
- 69 Alba AC, Agoritsas T, Walsh M, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *JAMA* 2017;318:1377–84.
- 70 Pencina MJ, Parikh CR, Kimmel PL, et al. Statistical methods for building better biomarkers of chronic kidney disease. *Stat Med* 2019;38:1903–17.
- 71 Hosmer DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. 3rd edn. Wiley, 2013.
- 72 Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994;69:979–85.
- 73 Whittle R, Royle KL, Jordan KP, et al. Prognosis research ideally should measure time-varying predictors at their intended moment of use. *Diagn Progn Res* 2017;1:1.
- 74 Mallett S, Royston P, Dutton S, et al. Reporting methods in studies developing prognostic models in cancer: a review. *BMC Med* 2010;8:20.
- 75 Pajouheshnia R, Groenwold RHH, Peelen LM, et al. When and how to use data from randomised trials to develop or validate prognostic models. *BMJ* 2019;365:l2154.



- 76 Eckardt K-U, Bärthlein B, Baid-Agrawal S, *et al.* The German Chronic Kidney Disease (GCKD) study: design and methods. *Nephrol Dial Transplant* 2012;27:1454–60.
- 77 Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health* 2001;22:189–212.
- 78 Benichou J, Gail MH. Methods of Inference for Estimates of Absolute Risk Derived from Population-Based Case-Control Studies. *Biometrics* 1995;51:182.
- 79 Heinze G, Wallisch C, Dunkler D. Variable selection – A review and recommendations for the practicing statistician. *Biometrical J* 2018;60:431–49.
- 80 Sauerbrei W, Perperoglou A, Schmid M, *et al.* State of the art in selection of variables and functional forms in multivariable analysis-outstanding issues. *Diagn Progn Res* 2020;4:3.
- 81 Goeman JJ, Solari A. Multiple hypothesis testing in genomics. *Statist Med* 2014;33:1946–78.
- 82 Ranstam J. Multiple P-values and Bonferroni correction. *Osteoarthritis Cartil* 2016;24:763–4.
- 83 Brinster R, Köttgen A, Tayo BO, *et al.* Control procedures and estimators of the false discovery rate and their application in low-dimensional settings: an empirical investigation. *BMC Bioinformatics* 2018;19:78.
- 84 Zeng ISL, Lumley T. Review of Statistical Learning Methods in Integrated Omics Studies (An Integrated Information Science). *Bioinform Biol Insights* 2018;12:1177932218759292.
- 85 Rahnenführer J, De Bin R, Benner A, *et al.* Statistical analysis of high-dimensional biomedical data: a gentle introduction to analytical goals, common approaches and challenges. *BMC Med* 2023;21:182.
- 86 Schumacher M, Holländer N, Schwarzer G, *et al.* Prognostic factor studies. In: *Handbook of statistics in clinical oncology*. 3rd edn. CRC Press Taylor & Francis, 2012.
- 87 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.
- 88 Machin D, ed. *Sample size tables for clinical studies*. 3rd edn. Wiley-Blackwell, 2008.
- 89 Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. *Stat Med* 1998;17:1623–34.
- 90 Schmoor C, Sauerbrei W, Schumacher M. Sample size considerations for the evaluation of prognostic factors in survival analysis. *Stat Med* 2000;19:441–52.
- 91 Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. *Biometrics* 1983;39:499–503.
- 92 Götte H, Zwiener I. Sample size planning for survival prediction with focus on high-dimensional data. *Stat Med* 2013;32:787–807.
- 93 Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? *JAMA* 2000;283:2701–11.
- 94 White IR, Kaptoge S, Royston P, *et al.* Meta-analysis of non-linear exposure-outcome relationships using individual participant data: A comparison of two methods. *Stat Med* 2019;38:326–38.
- 95 Sigounas DE, Tatsioni A, Christodoulou DK, *et al.* New prognostic markers for outcome of acute pancreatitis: overview of reporting in 184 studies. *Pancreas* 2011;40:522–32.
- 96 Sekula P, Pressler JB, Sauerbrei W, *et al.* Assessment of the extent of unpublished studies in prognostic factor research: a systematic review of p53 immunohistochemistry in bladder cancer as an example. *BMJ Open* 2016;6:e009972.
- 97 Peto R, Pike MC, Armitage P, *et al.* Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 1976;34:585–612.
- 98 Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ* 2016;352:i1981.
- 99 Peat G, Riley RD, Croft P, *et al.* Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. *PLoS Med* 2014;11:e1001671.
- 100 Mallett S, Timmer A, Sauerbrei W, *et al.* Reporting of prognostic studies of tumour markers: a review of published articles in relation to REMARK guidelines. *Br J Cancer* 2010;102:173–80.
- 101 Sekula P, Mallett S, Altman DG, *et al.* 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:e0178531.
- 102 McShane LM, Altman DG, Sauerbrei W, *et al.* REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;93:387–91.
- 103 Collins GS, Moons KGM, Dhiman P, *et al.* TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378.
- 104 Moons KGM, 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.
- 105 Altman DG. The time has come to register diagnostic and prognostic research. *Clin Chem* 2014;60:580–2.
- 106 Van Calster B, Wynants L, Riley RD, *et al.* Methodology over metrics: current scientific standards are a disservice to patients and society. *J Clin Epidemiol* 2021;138:219–26.
- 107 Sauerbrei W, Haeussler T, Balmford J, *et al.* Structured reporting to improve transparency of analyses in prognostic marker studies. *BMC Med* 2022;20:184.
- 108 Moher D, Glasziou P, Chalmers I, *et al.* Increasing value and reducing waste in biomedical research: who's listening? *Lancet* 2016;387:1573–86.
- 109 Thompson DR, Clark AM. The Ego Has Landed! What Can Be Done About Research Misconduct, Scandals, and Spins? *Am J Med* 2019;132:552–3.
- 110 Bradley SH, DeVito NJ, Lloyd KE, *et al.* Reducing bias and improving transparency in medical research: a critical overview of the problems, progress and suggested next steps. *J R Soc Med* 2020;113:433–43.
- 111 Keith DS, Nichols GA, Gullion CM, *et al.* Longitudinal follow-up and outcomes among a population with chronic kidney disease in a large managed care organization. *Arch Intern Med* 2004;164:659–63.
- 112 Govender MA, Brandenburg JT, Fabian J, *et al.* The Use of 'Omics for Diagnosing and Predicting Progression of Chronic Kidney Disease: A Scoping Review. *Front Genet* 2021;12:682929.