Prognosis and prognostic research: what, why, and how?

Karel G M Moons,[1] Patrick Royston,[2] Yvonne Vergouwe,[1] Diederick E Grobbee,[1] Douglas G Altman[3]

Doctors have little specific research to draw on when predicting outcome. In this first article in a series **Karel Moons and colleagues** explain why research into prognosis is important and how to design such research

[1]Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, Netherlands
[2]MRC Clinical Trials Unit, London NW1 2DA
[3]Centre for Statistics in Medicine, University of Oxford, Oxford OX2 6UD
Correspondence to: K G M Moons k.g.m.moons@umcutrecht.nl
Accepted: 6 October 2008

Cite this as: *BMJ* 2009;338:b375
doi: 10.1136/bmj.b375

This article is the first in a series of four aiming to provide an accessible overview of the principles and methods of prognostic research

Hippocrates included prognosis as a principal concept of medicine.[1] Nevertheless, principles and methods of prognostic research have received limited attention, especially compared with therapeutic and aetiological research. This article is the first in a series of four aiming to provide an accessible overview of these principles and methods. Our focus is on prognostic studies aimed at predicting outcomes from multiple variables rather than on studies investigating whether a single variable (such as a tumour or other biomarker) may be prognostic. Here we consider the principles of prognosis and multivariable prognostic studies and the reasons for and settings in which multivariable prognostic models are developed and used. The other articles in the series will focus on the development of multivariable prognostic models,[2] their validation,[3] and the application and impact of prognostic models in practice.[4]

Box 1 | **Consecutive phases in multivariable prognostic research**

- *Development studies*—Development of a multivariable prognostic model, including identification of the important predictors, assigning relative weights to each predictor, and estimating the model's predictive performance through calibration and discrimination and its potential for optimism using internal validation techniques, and, if necessary, adjusting the model for overfitting[2]
- *Validation studies*—Validating or testing the model's predictive performance (eg, calibration and discrimination) in new participants. This can be narrow (in participants from the same institution measured in the same manner by the same researchers though at a later time, or in another single institution by different researchers using perhaps slightly different definitions and data collection methods) or broad (participants obtained from various other institutions or using wider inclusion criteria)[3 4]
- *Impact studies*—Quantifying whether the use of a prognostic model by practising doctors truly improves their decision making and ultimately patient outcome, which can again be done narrowly or broadly.[4]

What is prognosis?

Prognosis simply means foreseeing, predicting, or estimating the probability or risk of future conditions; familiar examples are weather and economic forecasts. In medicine, prognosis commonly relates to the probability or risk of an individual developing a particular state of health (an outcome) over a specific time, based on his or her clinical and non-clinical profile. Outcomes are often specific events, such as death or complications, but they may also be quantities, such as disease progression, (changes in) pain, or quality of life.

In medical textbooks, however, prognosis commonly refers to the expected course of an illness. This terminology is too general and has limited utility in practice. Doctors do not predict the course of an illness but the course of an illness in a particular individual. Prognosis may be shaped by a patient's age, sex, history, symptoms, signs, and other test results. Moreover, prognostication in medicine is not limited to those who are ill. Healthcare professionals, especially primary care doctors, regularly predict the future in healthy individuals—for example, using the Apgar score to determine the prognosis of newborns, cardiovascular risk profiles to predict heart disease in the general population, and prenatal testing to assess the risk that a pregnant woman will give birth to a baby with Down's syndrome.

Multivariable research

Given the variability among patients and in the aetiology, presentation, and treatment of diseases and other health states, a single predictor or variable rarely gives an adequate estimate of prognosis. Doctors—implicitly or explicitly—use multiple predictors to estimate a patient's prognosis. Prognostic studies therefore need to use a multivariable approach in design and analysis to determine the important predictors of the studied outcomes and to provide outcome probabilities for different combinations of predictors, or to provide tools to estimate such probabilities. These tools are commonly called prognostic models, prediction models, prediction rules, or risk scores.[5-14] They enable care providers to use combinations of predictor values to estimate

an absolute risk or probability that an outcome will occur in an individual. A multivariable approach also enables researchers to investigate whether specific prognostic factors or markers that are, say, more invasive or costly to measure, have worthwhile added predictive value beyond cheap or simply obtained predictors—for example, from patient history or physical examination. Nonetheless, many prognostic studies still consider a single rather than multiple predictors.[15]

Use of prognostic models

Medical prognostication and prognostic models are used in various settings and for various reasons. The main reasons are to inform individuals about the future course of their illness (or their risk of developing illness) and to guide doctors and patients in joint decisions on further treatment, if any. For example, modifications of the Framingham cardiovascular risk score[16] are widely used in primary care to determine the indication for cholesterol lowering and antihypertensive drugs. Examples from secondary care include use of the Nottingham prognostic index to estimate the long term risk of cancer recurrence or death in breast cancer patients,[17] the acute physiology and chronic health evaluation (APACHE) score and simplified acute physiology score (SAPS) to predict hospital mortality in critically ill patients,[18 19] and models for predicting postoperative nausea and vomiting.[20 21]

Another reason for prognostication and use of prognostic models is to select relevant patients for therapeutic research. For example, researchers used a previously validated prognostic model to select women with an increased risk of developing cancer for a randomised trial of tamoxifen to prevent breast cancer.[22] Another randomised trial on the efficacy of radiotherapy after breast conserving resection used a prognostic model to select patients with a low risk of cancer recurrence.[23]

Prognostic models are also used to compare differences in performance between hospitals. For example, the clinical risk index for babies (CRIB) was originally developed to compare performance and mortality among neonatal intensive care units.[24] More recently Jarman et al developed a model to predict the hospital standardised mortality ratio to explain differences between English hospitals.[25]

Differences from aetiological research

Although there are clear similarities in the design and analysis of prognostic and aetiological studies, predicting outcomes is not synonymous with explaining their cause.[26 27] In aetiological research, the mission is to explain whether an outcome can reliably be attributed to a particular risk factor, with adjustment for other causal factors (confounders) using a multivariable approach. In prognostic research, the mission is to use multiple variables to predict, as accurately as possible, the risk of future outcomes. Although a prognostic model may be used to provide insight into causality or pathophysiology of the studied outcome, that is neither an aim nor a requirement. All variables potentially associated with the outcome, not necessarily causally, can be considered in a prognostic study. Every causal factor is a predictor—albeit sometimes a weak one—but not every predictor is a cause. Nice examples of predictive but non-causal factors used in everyday practice are skin colour in the Apgar score and tumour markers as predictors of cancer progression or recurrence. Both are surrogates for obvious causal factors that are more difficult to measure.

Furthermore, to guide prognostication in individuals, analysis and reporting of prognostic studies should focus on absolute risk estimates of outcomes given combinations of predictor values. Relative risk estimates (for example, odds ratio, risk ratio, or hazard ratio) have no direct meaning or relevance to prognostication in practice. In prediction research, relative risks are used only to obtain an absolute probability of the outcome for an individual, as we will show in our second article.[2] In contrast, aetiological and therapeutic studies commonly focus on relative risks—for example, the risk of an outcome in presence of a causal factor relative to the risk in its absence. Also, the calibration and discrimination of a multivariable model are highly relevant to prognostic research but meaningless in aetiological research.

How to study prognosis?

Building on previous guidelines[8 10 14 28 29] we distinguish three major steps in multivariable prognostic research that are also followed in the other articles in this series[2-4]: developing the prognostic model, validating its performance in new patients, and studying its clinical impact (box). We focus here on the non-statistical characteristics of a multivariable study aimed at developing a prognostic model. The statistical aspects of developing a model are covered in our second article.[2]

Objective

The main objective of a prognostic study is to determine the probability of the specified outcome with different combinations of predictors in a well defined population.

Study sample

The study sample includes people at risk of developing the outcome of interest, defined by the presence of a particular condition (for example, an illness, undergoing surgery, or being pregnant).

Study design

The best design to answer prognostic questions is a cohort study. A prospective study is preferable as it enables optimal measurement of predictors and outcome (see below). Studies using cohorts already assembled for other reasons allow longer follow-up times but usually at the expense of poorer data. Unfortunately, the prognostic literature is dominated by retrospective studies. Case-control studies are sometimes used for prognostic analysis, but they do not automatically allow estimation of absolute risks because cases and controls are often sampled from a source population of unknown size. Since investigators are free to choose the ratio of

cases and controls, the absolute outcome risks can be manipulated.[30] An exception is a case-control study nested in a cohort of known size.[31]

Data from randomised trials of treatment can also be used to study prognosis. When the treatment is ineffective (relative risk=1.0), the intervention and comparison group can simply be combined to study baseline prognosis. If the treatment is effective the groups can be combined, but the treatment variable should then be included as a separate predictor in the multivariable model. Here treatments are studied on their independent predictive effect and not on their therapeutic or preventive effects. However, prognostic models obtained from randomised trial data may have restricted generalisability because of strict eligibility criteria for the trial, low recruitment levels, or large numbers refusing consent.

Predictors

Candidate predictors can be obtained from patient demographics, clinical history, physical examination, disease characteristics, test results, and previous treatment. Prognostic studies may focus on a cohort of patients who have not (yet) received prognosis modifying treatments—that is, to study the natural course or baseline prognosis of patients with that condition. They can also examine predictors of prognosis in patients who have received treatments.

Studied predictors should be clearly defined, standardised, and reproducible to enhance generalisability and application of study results to practice.[32] Predictors requiring subjective interpretation, such as imaging test results, are of particular concern in this context because there is a risk of studying the predictive ability of the observer rather than that of the predictors. Also, predictors should be measured using methods applicable—or potentially applicable—to daily practice. Specialised measurement techniques may yield optimistic predictions.

As discussed above, the prognostic value of treatments can also be studied, especially when randomised trials are used. However, caution is needed in including treatments as prognostic factors when data are observational. Indications for treatment and treatment administration are often not standardised in observational studies and confounding by indication could lead to bias and large variation in the (type of) administered treatments.[33] Moreover, in many circumstances the predictive effect of treatments is small compared with that of other important prognostic variables such as age, sex, and disease stage.

Finally, of course, studies should include only predictors that will be available at the time when the model is intended to be used.[34] If the aim is to predict a patient's prognosis at the time of diagnosis, for example, predictors that will not be known until actual treatment has started are of little value.

Outcome

Preferably, prognostic studies should focus on outcomes that are relevant to patients, such as occurrence or remission of disease, death, complications, tumour growth, pain, treatment response, or quality of life. Surrogate or intermediate outcomes, such as hospital stay or physiological measurements, are unhelpful unless they have a clear causal relation to relevant patient outcomes, such as CD4 counts instead of development of AIDS or death in HIV studies. The period over which the outcome is studied and the methods of measurement should be clearly defined. Finally, outcomes should be measured without knowledge of the predictors under study to prevent bias, particularly if measurement requires observer interpretation. Blinding is not necessary when the outcome is all cause mortality. But, if the outcome is cause specific mortality, knowledge of the predictors might influence assessment of outcomes (and vice versa in retrospective studies where predictors are documented after the outcome was assessed).

Required number of patients

The multivariable character of prognostic research makes it difficult to estimate the required sample size. There are no straightforward methods for this. When the number of predictors is much larger than the number of outcome events, there is a risk of overestimating the predictive performance of the model. Ideally, prognostic studies require at least several hundred outcome events. Various studies have suggested that for each candidate predictor studied at least 10 events are required,[6 8 35 36] although a recent study showed that this number could be lower in certain circumstances.[37]

Validation and application of prognostic models

Formally developed and validated prognostic models are often used in weather forecasting and economics (with varying success), but not in medicine. There may be several reasons for this. Firstly, prognostic models are often too complex for daily use in clinical settings without computer support. The introduction of computerised patient records will clearly enhance not only the development and validation of models in research settings but also facilitate their application in routine care.[38 39] Secondly, because many prognostic models have not been validated in other populations, clinicians may (and perhaps should) not trust probabilities provided by these models.[14 40-42]

Finally, clinicians often do not know how to use predicted probabilities in their decision making. Validation studies are scarce, but even fewer models are tested for their ability to change clinicians' decisions, let alone to change patient outcome.[14] We support the view that no prediction model should be implemented in practice until, at a minimum, its performance has been validated in new individuals.[6-10 12 14 29 43 44] The third article in this series discusses why validation studies are important and how to design and interpret them.[3]

Validation studies are particularly important if a prediction model is to be used in individuals who were not represented in the development study—for example, when transporting a model from secondary to primary care or from adults to children, which seems a form of extrapolation rather than validation.[43 45] We will

SUMMARY POINTS

Prognosis is estimating the risk of future outcomes in individuals based on their clinical and non-clinical characteristics

Predicting outcomes is not synonymous with explaining their cause

Prognostic studies require a multivariable approach to design and analysis

The best design to address prognostic questions is a cohort study

discuss this further in the fourth article in the series, as well as how to update existing models to other circumstances.[4]

We stress that prediction models are not meant to take over the job of the doctor.[7 40 41 46] They are intended to help doctors make decisions by providing more objective estimates of probability as a supplement to other relevant clinical information. Furthermore, they improve understanding of the determinants of the course and outcome of patients with a particular disease.

Funding: KGMM, YV, and DEG are supported by the Netherlands Organisation for Scientific Research (ZON-MW 917.46.360). PR is supported by the UK Medical Research Council (U.1228.06.001.00002.01). DGA is supported by Cancer Research UK.

Contributors: The four articles in the series were conceived and planned by DGA, KGMM, PR, and YV. KGMM wrote the first draft of this article. All the authors contributed to subsequent revisions.

Competing interests: None declared.

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

1 Hippocrates. On airs, waters and places. In: Adams F, ed. *The genuine works of Hippocrates*. Baltimore: Wilkins and Wilkins, 1939.
2 Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009;338:b604.
3 Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: Validating a prognostic model. *BMJ* 2009;338:b605.
4 Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: Application and impact of prognostic models in clinical practice. *BMJ* 2009:338:b606.
5 Laupacis A, Wells G, Richardson WS, Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. *JAMA* 1994;272:234-7.
6 Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87.
7 Braitman LE, Davidoff F. Predicting clinical states in individual patients. *Ann Intern Med* 1996;125:406-12.
8 Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;277:488-94.
9 Randolph AG, Guyatt GH, Calvin JE, Doig DVM, Richardson WS. Understanding articles describing clinical prediction tools. *Crit Care Med* 1998;26:1603-12.
10 Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19:453-73.
11 Concato J. Challenges in prognostic analysis. *Cancer* 2001;91:1607-14.
12 Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med* 2004;23:2567-86.
13 McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumour marker prognostic studies (REMARK). *Br J Cancer* 2005;93:387-91.
14 Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006;144:201-9.
15 Riley RD, Abrams KR, Sutton AJ, Lambert PC, Jones DR, Heney D, et al. Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. *Br J Cancer* 2003;88:1191-8.
16 Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham study. *Am J Cardiol* 1976;38:46-51.
17 Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham prognostic index in primary breast cancer. *Breast Cancer Res Treat* 1992;22:207-19.
18 Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619-36.
19 Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993;270:2957-63.
20 Van den Bosch JE, Moons KG, Bonsel GJ, Kalkman CJ. Does measurement of preoperative anxiety have added value for predicting postoperative nausea and vomiting? *Anesth Analg* 2005;100:1525-32.
21 Apfel CC, Laara E, Koivuranta M, Greim CA, Roewer N. A simplified risk score for predicting postoperative nausea and vomiting: conclusions from cross-validations between two centers. *Anesthesiol* 1999;91:693-700.
22 Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for prevention of breast cancer: report of the national surgical adjuvant breast and bowel project P-1 study. *J Natl Cancer Inst* 1998;90:1371-88.
23 Winzer KJ, Sauer R, Sauerbrei W, Schneller E, Jaeger W, Braun M, et al. Radiation therapy after breast-conserving surgery; first results of a randomised clinical trial in patients with low risk of recurrence. *Eur J Cancer* 2004;40:998-1005.
24 Cockburn F CR, Gamsu HR, Greenough A, Hopkins A, International Neonatal Network. The CRIB (clinical risk index for babies) score: a tool assessing initial neonatal risk and comparing performance of neonatal intensive care units. *Lancet* 1993;342:193-8.
25 Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999;318:1515-20.
26 Moons KG, Grobbee DE. Clinical epidemiology: an introduction. In: Vaccaro AR, ed. *Orthopedic Knowledge Update: 8*. Rosemont: American Academy of Orthopaedic Surgeons, 2005:109-18.
27 Brotman DJ, Walker E, Lauer MS, O'Brien RG. In search of fewer independent risk factors. *Arch Intern Med* 2005;165:138-45.
28 Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 1985;313:793-9.
29 McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. *JAMA* 2000;284:79-84.
30 Moons KG, van Klei W, Kalkman CJ. Preoperative risk factors of intraoperative hypothermia in major surgery under general anesthesia. *Anesth Analg* 2003;96:1843-4.
31 Biesheuvel CJ, Vergouwe Y, Oudega R, Hoes AW, Grobbee DE, Moons KG. Rehabilitation of the nested case-control design in diagnostic research. *BMC Med Res Methodol* 2008;8:48.
32 Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994;69:979-85.
33 Grobbee DE. Confounding and indication for treatment in evaluation of drug treatment for hypertension. *BMJ* 1997;315:1151-4.
34 Walraven vC, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. *J Clin Epidemiol* 2004;57:672-82.
35 Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. *J Clin Epidemiol* 1995;48:1495-501.
36 Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995;48:1503-10.
37 Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007;165:710-8.
38 James BC. Making it easy to do it right. *N Engl J Med* 2001;345:991-3.
39 Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 2005;330:765.
40 Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med* 1993;118:201-10.
41 Christensen E. Prognostic models including the Child-Pugh, MELD and Mayo risk scores—where are we and where should we go? *J Hepatol* 2004;41:344-50.
42 Wyatt JC, Altman DG. Prognostic models: clinical useful or quickly forgotten? *BMJ* 1995;311:1539-41.
43 Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130:515-24.
44 Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;56:826-32.
45 Knottnerus JA. Between iatrotropic stimulus and interiatric referral: the domain of primary care research. *J Clin Epidemiol* 2002;55:1201-6.
46 Feinstein AR. Clinical Judgment revisited: the distraction of quantitative models. *Ann Intern Med* 1994;120:799-805.