

Review

Prognostic markers in cancer: the evolution of evidence from single studies to meta-analysis, and beyond

RD Riley^{*1}, W Sauerbrei² and DG Altman³

¹Department of Public Health, Epidemiology and Biostatistics, Public Health Building, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK; ²Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Stefan-Meier-Strasse 26, Freiburg 79104, Germany; ³Centre for Statistics in Medicine, University of Oxford, Wolfson College Annexe, Linton Road, Oxford OX2 6UD UK

In oncology, prognostic markers are clinical measures used to help elicit an individual patient's risk of a future outcome, such as recurrence of disease after primary treatment. They thus facilitate individual treatment choice and aid in patient counselling. Evidence-based results regarding prognostic markers are therefore very important to both clinicians and their patients. However, there is increasing awareness that prognostic marker studies have been neglected in the drive to improve medical research. Large protocol-driven, prospective studies are the ideal, with appropriate statistical analysis and clear, unbiased reporting of the methods used and the results obtained. Unfortunately, published prognostic studies rarely meet such standards, and systematic reviews and meta-analyses are often only able to draw attention to the paucity of good-quality evidence. We discuss how better-quality prognostic marker evidence can evolve over time from initial exploratory studies, to large protocol-driven primary studies, and then to meta-analysis or even beyond, to large prospectively planned pooled analyses and to the initiation of tumour banks. We highlight articles that facilitate each stage of this process, and that promote current guidelines aimed at improving the design, analysis, and reporting of prognostic marker research. We also outline why collaborative, multi-centre, and multi-disciplinary teams should be an essential part of future studies.

British Journal of Cancer (2009) **100**, 1219–1229. doi:10.1038/sj.bjc.6604999 www.bjcancer.com
© 2009 Cancer Research UK

Keywords: prognosis; prognostic tumour marker; biomarker; primary studies; evidence synthesis; guidelines

In oncology, prognostic markers (also called prognostic factors) are clinical measures used to help elicit an individual patient's risk of a future outcome, such as recurrence of disease after primary treatment. They play a key role in clinical practice, distinguishing patients into different risk groups and thus informing treatment strategies and aiding patient counselling. They can also be used to define strata in clinical trials to ensure comparability of treatment groups. Markers can be simple measures, such as the stage of disease or tumour size, but are often more complex, such as abnormal levels of proteins or genetic mutations. For example, within paediatric oncology, amplification of the MYCN proto-oncogene is a known indicator of poor outcome in neuroblastoma patients (Riley *et al*, 2004a). In adult oncology, it is well known that the number of positive lymph nodes has a strong influence on the prognosis of recurrence-free survival time in patients with primary breast cancer (Galea *et al*, 1992). For (nearly) all diseases, some important prognostic markers have been well established for a long time, but many more are investigated each year. Few, however, find a role in clinical practice.

Not all potential markers turn out to have prognostic value. Much research effort goes into conducting studies that evaluate the extent to which certain markers help clinical prognosis. Indeed, hundreds of prognostic marker studies are published in cancer

journals each year. Unfortunately, though, there is large concern about the quality of such studies and it is clear that large progress is needed to produce clinically relevant results (Hinestrosa *et al*, 2007). Over more than 10 years, an increasing body of evidence has signalled that prognostic marker studies are often badly designed (Simon and Altman, 1994; Altman and Lyman, 1998), inappropriately analysed (Holländer and Sauerbrei, 2006), poorly reported (Riley *et al*, 2003; Kyzas *et al*, 2005), and subject to numerous biases, such as selective reporting (Kyzas *et al*, 2005, 2007a) and 'optimal' choice of cutpoints (Altman *et al*, 1994). Bad quality of design, analysis, and reporting of individual studies result in confusion regarding the prognostic value of a new marker (Sauerbrei, 2005). Furthermore, evidence-based marker results are a rarity, with systematic reviews and meta-analyses of multiple studies serving only to highlight the serious deficiencies within primary prognostic research (Altman, 2001). For example, two systematic reviews, one in neuroblastoma (Riley *et al*, 2004a) and another in non-small-cell-lung cancer (Brundage *et al*, 2002), found that over 100 different prognostic markers have been investigated in each field, yet the median number of articles per marker was only 1. Clinicians are thus faced with a confusing yet ever-growing body of literature, where new markers are regularly investigated but rarely in relation to existing markers. The evidence-base is further blurred by large heterogeneity across studies in the patient population, the method of measuring each marker, the outcomes reported, and the statistical analyses used, among numerous other factors (Riley *et al*, 2003), all of which limit a coherent meta-analysis.

*Correspondence: Dr RD Riley; E-mail: r.d.riley@bham.ac.uk

Received 13 November 2008; revised 4 February 2009; accepted 4 March 2009

In this paper, we examine the key methodological issues of the prognostic marker research process, encouraging a move towards higher standards and clinically useful results. Our aim is to discuss how high-quality evidence can evolve over time from initial exploratory studies, to large protocol-driven primary studies, to meta-analysis and even beyond, to large prospectively planned pooled analyses, and to the initiation of tissue banks. The work extends earlier articles aimed at a more statistical audience (Altman *et al*, 2006; Holländer and Sauerbrei, 2006; Sauerbrei *et al*, 2006; Riley *et al*, 2006a), and focuses on prognostic, rather than predictive, markers. In cancer, a prognostic marker is one that predicts a patient's clinical course, whereas a predictive marker is one associated with differential responses to a specific treatment (although this terminological distinction is perhaps not widespread across medicine). Predictive marker studies are relatively uncommon, and best embedded within randomised controlled trials. Sargent *et al* (2005) discuss how to conduct a clinical trial that assesses the utility of a predictive marker.

Our paper focuses on studies (or meta-analyses) that seek to evaluate a small number of pre-specified markers. In recent years, clinicians have become increasingly enthusiastic about the possibility of improving prognosis and prediction by applying the new technologies of measuring many thousands of genes. However, these technologies create a very large number of potential predictors and require statistical methods to extract a relatively small number of relevant items. We also do not consider the study of prognostic markers from high-dimensional data, such as microarray studies (Tinker *et al*, 2006; Dupuy and Simon, 2007).

The outline of the paper is as follows. In the section 'Primary studies of prognostic markers', we consider primary studies of prognostic markers and consider their design, analysis, and reporting. In the section 'Systematic reviews and meta-analysis', we discuss systematic reviews of such studies and explain why meta-analyses that use individual patient data (IPD) are important. In the section 'Towards multi-centre and collaborative research', we consider why prospectively planned pooled analyses and the initiation of tumour banks may best facilitate evidence-based prognostic marker results. Finally, in the Discussion section, we summarise the key messages for future research.

PRIMARY STUDIES OF PROGNOSTIC MARKERS

Study design

In the initiation of prognostic marker studies, as Altman and Lyman (1998) suggest, every effort must be made to limit potential biases and emulate the design standards expected of a protocol-driven randomised trial. Researchers should begin with a clear and well-defined research question so that their study can be designed and targeted appropriately. Unfortunately, it seems that many prognostic studies are not motivated by prior questions or by hypotheses, and few are protocol-driven – 'investigators may tend to conduct opportunistic studies on the basis of specimen availability rather than on thoughtful design' (Kyzas *et al*, 2007b). For

example, often no justification is made for the sample size used, or whether the markers, sub-populations, and outcomes assessed were determined before beginning the study. It can be speculated that the main motivation for a quick analysis of some pre-existing data is another publication rather than scientific importance. It is, however, good practice for all research to have a study protocol outlining the aims and detailing the methods that will be used.

Prognostic marker studies have several possible aims, including the desire to understand more about the biology of a disease process. However, the vast majority of studies aim to investigate the association between one or more markers with an outcome (or multiple outcomes) of interest. In cancer, these outcomes are usually death, recurrence of disease, or both. For such studies, three different study phases have been proposed (Figure 1), which are similar in spirit to those proposed for biomarker development (Pepe *et al*, 2001; Ransohoff, 2007), and prognostic factor studies in general (Hayden *et al*, 2008). They begin with exploratory studies to identify promising prognostic markers and move towards larger, preferably prospective, confirmatory studies. Phase I studies are early exploratory analyses to generate hypotheses and to identify potential markers for further investigation. Phase II studies continue the exploratory investigation and assess the relationship between marker and prognosis. Phase III studies are large, confirmatory studies that state prior hypotheses, potentially based on earlier phase II results. These studies should certainly be protocol-driven and must be seen as the highest level of evidence that can arise from an individual prognostic study.

It is important for researchers to consider *a priori* which of these study types fit their research objective. There is nothing wrong with exploratory and hypothesis generating phase I or II studies. Furthermore, additional exploratory analyses of phase III studies are encouraged, as all such exploratory works help to identify potentially important markers and to validate measurement techniques. However, it is important that this type of research is clearly labelled and interpreted as such. After the publication of several small phase I or II studies, it seems useless, or even harmful, to proceed with further such studies for a particular marker of interest (Schmitz-Dräger *et al*, 2000). Research efforts should rather concentrate on larger collaborative phase III projects. The availability of a study register would help researchers to decide about starting further smaller studies or whether the time has come for a collaborative phase III study. Study registers would also prevent duplication of research effort, and allow transparency of *all* the markers that have been investigated, not just those found to be potentially important. Study registers are an integral part of randomised trials (Horton and Smith, 1999), thanks to many years of petitioning (Simes, 1986), and this sets the standard for transparent prognostic marker studies to follow. We note that such registers may be easier to achieve for phase III prognostic studies, which are pre-planned and protocol-driven by nature. On the contrary, a concern is that, owing to their opportunistic nature, many phase I or II studies would only be registered if they identify a significant finding, resulting in a biased set of registered phase I and II studies, akin to the problem of publication bias. Registration before the start of a study would be ideal, but we are aware that it

1. Phase I: Exploratory studies (hypothesis generating) that seek an association between a prognostic marker and characteristics of disease thought to have prognostic importance.
2. Phase II: Exploratory studies attempting to use values of a prognostic marker to discriminate between patients at high and low risk of disease progression or death.
3. Phase III: Confirmatory studies of *a priori* hypotheses attempting to use values of a prognostic marker to discriminate between patients at high and low risk of disease progression or death.

Figure 1 Types of prognostic marker studies, modified from Altman and Lyman (1998).

will be difficult to achieve such a situation with phase I or II studies.

Unfortunately, the current literature indicates that phase I and II studies are by far the most common, with the higher-quality phase III studies unfortunately quite rare. This is disappointing, as results from phase III studies are more reliable and should underpin which markers are implemented in practice. Simon and Altman (1994) presented requirements of a prognostic marker for acceptance in clinical practice (Figure 2), and these include confirmation from phase III studies that the marker has independent prognostic ability beyond other markers used in practice (Kattan, 2003). Factors to consider in the design of phase III studies are shown in Figure 3. An ideal approach would involve following a well-defined cohort of patients from the same stage of their disease. Such a sample is often an 'inception' cohort of patients at the time of diagnosis of cancer. The use of study protocols, pre-specified study objectives, and prior hypotheses are essential. Detailed consideration should be given to the sample selection, inclusion and exclusion criteria, treatment (if possible randomised or standardised), the use of standardised assays, and statistical analysis (see later). For phase III studies, sample size should be large enough to sustain meaningful analysis, preferably with hundreds of events, and to ensure that the marker of interest can be assessed alongside existing markers of prognostic importance. Kyzas *et al* (2007b) highlight that power calculations for the required sample size are rarely presented in published prognostic marker studies. Procedures are also needed to ensure missing data are minimised and completeness of follow-up is maximised.

- Determination is reproducible and widely available with quality control.
- Substantial added value beyond recognised prognostic systems is shown.
- Conclusions are based on independent confirmed phase III studies.

Figure 2 Requirements of a prognostic marker for acceptance in clinical practice, modified from Simon and Altman (1994).

Although protocol-driven prospective studies are the ideal (Van Meerbeeck, 1994; Altman and Lyman, 1998), unfortunately the large majority of prognostic studies in cancer are retrospective and, it seems, not protocol-driven. A review of 331 prognostic marker studies identified only 20% that were prospective (Kyzas *et al*, 2007b), in the sense that the data were collected after the research question was posed. The big advantage of retrospective studies is the availability of a cohort with a long enough follow-up to assess a substantial number of outcome events (deaths or recurrences). Retrospective studies, however, have several serious disadvantages. Foremost among these are problems associated with the lack of a fully specified design for the study: unclear inclusion criteria, unknown completeness of the cohort, lack of standardisation of diagnostic and therapeutic procedures, incomplete baseline data, and unclear completeness of follow-up. A move towards prospective prognostic studies is thus encouraged, especially within phase III research. This may require collaboration among multiple research groups, so to achieve larger sample sizes and consistency across groups in important factors (e.g., measurement techniques, cutpoints used). Such studies can best be integrated in large therapeutic trials, but a protocol-based pre-planned pooled analysis of retrospective data from several centres is also possible (see the section 'Towards multi-centre and collaborative research').

Statistical analysis

For therapeutic studies, statistical principles and methods are well developed and generally accepted. In contrast, no such consensus exists for the evaluation of prognostic markers. Holländer and Sauerbrei (2006) showed that the choice of statistical method has a strong influence on the results and, therefore, on the interpretation of prognostic marker studies. Focusing on prognostic markers in oncology, they discuss issues of statistical model building in the framework of regression models and of classification and regression trees. Other approaches, for example, artificial neural

- The primary and secondary hypotheses should be clearly stated, including any subgroup analyses in advance of the study.
- Consider prognostic markers for which there is sufficient evidence to warrant further investigation based on:
 - (a) previous studies;
 - (b) biological and clinical plausibility;
 - (c) relevance and importance to the understanding or treatment of the disease.
- The study population should be defined with specified inclusion and exclusion criteria and methods to judge evaluability.
- Patient treatment should if possible be either rule-based (several treatment strategies are allowed) or assigned by randomisation.
- Assays should be reproducible and should be performed without knowledge of the clinical data and patient outcome. Preferable to use a standardised assay that is commonly used and available.
- Estimate the sample size keeping in mind the following:
 - (a) the desired power to detect meaningful difference for the major end points and to reject such differences with reasonable confidence if they are not found;
 - (b) the relationship of sample size to the number of outcome events, bearing in mind that these will be less frequent in favourable prognostic groups;
 - (c) the desirability of large prospective studies of a single prognostic marker.
- Outline how missing data will be minimised, and completeness of patient follow-up maximised.
- Specify the statistical analysis plan (SAP) including details on cutpoints or the modelling of functional relationship for continuous variables, proposed hypothesis testing on subgroups, anticipated interactions in advance of the study, and how missing data will be handled.
- Key study features, including the above information, should be fully detailed in a formal written protocol.

Figure 3 Factors to consider in the design of a high-quality phase III study, modified from Altman and Lyman (1998).

networks, are sometimes used for the analysis of marker studies but have several drawbacks (for more detailed discussion on this topic see Schwarzer *et al*, 2000; Schumacher *et al*, 2006). Parametric survival models are also possible, but are rarely used for the identification of prognostic markers.

For the analysis of a prognostic marker study, we consider multivariable regression models as the method of choice, such as the Cox model, which is often suitable for survival-type data that arises from prognostic studies. This assessment is based on both personal experiences and the results of simulation studies considering a restricted range of model-building issues (Sauerbrei, 1999; Royston and Sauerbrei, 2008). Multivariate analyses are easy to perform and provide prognostic marker results adjusted for other markers; this is imperative to assess the added value of a new marker over existing markers (Kattan, 2003), which should be the standard practice but is currently the exception in the prognostic marker literature (Kyzas *et al*, 2005).

Within regression models, a central issue is whether to use the full model incorporating all available variables or a reduced model determined by variable selection methods. Although it is well known that variable selection has several difficulties (Sauerbrei, 1999; Harrell, 2001), these strategies are often required to derive a sensible, interpretable, and parsimonious model (Sauerbrei, 1999; Royston and Sauerbrei, 2008). A more crucial issue is the way in which continuous markers are analysed. We believe it is central to make maximal use of the data and determine a sensible functional form (Royston *et al*, 2006; Sauerbrei *et al*, 2007; Royston and Sauerbrei, 2008). Researchers may retain a variable as a continuous measurement but they often assume a linear relation to outcome, which is unwise. More often they choose to avoid such issues by dichotomising, often at the sample median. This popular 'step function' approach contradicts biological thinking and has major methodological deficiencies, not least the reduced statistical power (Royston *et al*, 2006). Using the so-called 'optimal cutpoint' approach is the worst form of the step function approach, as it introduces considerable bias (Altman *et al*, 1994; Royston *et al*, 2006). Use of more than one cutpoint is better than having just one, but we believe that full use of the available data mandates a careful analysis that uses the actual continuous marker values (Holländer and Sauerbrei, 2006; Royston and Sauerbrei, 2008). Recommendations towards model building by the selection of variables and functional forms for continuous markers are available (Sauerbrei *et al*, 2007).

An important statistical topic that is often neglected in prognostic marker studies is model validation with external data (Holländer and Sauerbrei, 2006). Schumacher *et al* (2006) state that: 'A multivariate approach is absolutely essential. Thoughtful application of model building techniques should help to obtain models that are as simple and parsimonious as possible and to avoid serious overfitting in order to achieve generalisability for future patients. Thus, validation in an independent study is a further essential step'. Unfortunately, most markers or combination of markers used as the basis for classification schemes never undergo external validation in a new data, and those that do rarely maintain their prognostic ability (Sauerbrei *et al*, 1997); consequently, many so-called prognostic markers are often not accepted for general use (Wyatt and Altman, 1995; Boracchi and Biganzoli, 2003). Altman and Royston (2000) propose that statistical and clinical validation of prognostic models is required, and they examine some general approaches to do this. One way of validating whether a marker truly is prognostic is to look at the results across multiple studies (meta-analysis), but this may be difficult without IPD (see section 'The benefit of IPD') (Tudur-Smith *et al*, 2005). Bennett (2003) provides a helpful review of analytic methods for analysing time-to-event data in single studies and meta-analysis, considering non-standard issues such as non-proportional hazards and heterogeneity.

Kattan (2003), and also Kozt and Kattan (2005), highlights that the added value of new prognostic markers should be assessed not

only by multivariate analyses but also through their ability to improve predictive accuracy. Two new measures applicable to models for binary outcome are proposed by Pencina *et al* (2008). Pepe *et al* (2004) also state that: 'Markers proposed for classifying or predicting risk in individual subjects must be held to a much higher standard than merely being associated with outcome.' Vickers and Elkin (2006) provide a novel method for evaluating prediction models, whereas Henderson and Keiding (2005) argue that although prediction models can be useful at the group or the population level, they are usually of no real use for individual patients because of the large uncertainty in human survival.

In Figure 4, we reproduce the important issues for the analysis of a prognostic marker study from Holländer and Sauerbrei (2006). As for all high-quality, protocol-driven research, we recommend that the statistical analysis plan of a prognostic marker study should be specified in the study protocol and include details of how the independent prognostic importance of the marker will be ascertained in relation to existing markers of clinical importance. Furthermore, such pre-specified analyses should be complemented by sensitivity analyses of central assumptions and by additional model building using suitable strategies. These latter results are explanatory, which should be clarified in the study report.

Study reporting

In addition to improving the design and analysis of primary studies, there is also a pressing need to improve their reporting standards, as serious deficiencies have been exposed (Altman *et al*, 1995; Riley *et al*, 2003; Burton and Altman, 2004; Kyzas *et al*, 2005, 2007a). Published studies currently lack sufficient information to allow a full appreciation of the methodological quality of the study, the methods, and analyses undertaken, or the applicability of the study results for practice. For instance, the following rudimentary factors are often not reported: the number of patients in each marker group; the number of events in each group; the method of measuring each marker; the cutpoint used to dichotomise a continuous marker into 'high' and 'low' levels; a measure of effect, such as a hazard ratio, and its confidence interval; marker results adjusted for other clinically useful markers; and which analyses, outcome, and markers were primary (pre-defined) objectives and which were exploratory secondary assessments (Riley *et al*, 2003). Kyzas *et al* (2007b) reviewed 331 cancer prognostic studies from 20 meta-analyses, and concluded that the reporting quality of study design and assay information was often suboptimal. Such gaps in reporting may explain why most systematic reviews of prognostic markers do not include an adequate quality appraisal of the studies identified (Hayden *et al*, 2006).

There is also evidence that prognostic results are often subject to selective reporting, meaning that some of the markers, outcomes, and analyses considered are not reported upon publication (Kyzas *et al*, 2005). This issue is associated with the common threat of publication bias (Sterne *et al*, 2001; Rothstein *et al*, 2005), where studies that do not identify statistically or clinically significant results are not published. Such publication bias is well recognised for randomised trials, even for Food and Drug Administration (FDA)-registered studies (Turner *et al*, 2008), and it seems certain that this bias also affects prognostic studies to a great extent. The biggest example so far was shown by Kyzas *et al* (2007a), who found that <1.5% of 1915 articles on cancer prognostic markers were fully 'negative,' in that they did not present statistically significant prognostic results and did not, for example, elaborate on nonsignificant trends. Selective publication is likely to lead to larger effects that are seen in smaller studies (Sterne *et al*, 2001; Rothstein *et al*, 2005). For example, in a systematic review of studies of Bcl2 in non-small-cell lung cancer, almost all the smaller studies showed a statistically significant relationship between Bcl2 and the risk of dying, whereas the three large studies were all nonsignificant and showed a much smaller effect (Martin *et al*,

1. *Treatment of study population:*
 - ideally patients without any systemic adjuvant therapy, but unrealistic in many diseases;
 - standardisation or randomisation of treatment preferable, adjustment for treatment in the analysis.
2. *Model-building process:*
 - multivariable model required to assess the effect of a marker;
 - many approaches available, no general agreement concerning preferable strategies;
 - our preference are regression models, other approaches can give complementary information.
 - 2a Regression models:
 - problems caused by categorisation of continuous covariates;
 - standard assumption of a (log-)linear effect of continuous markers may be wrong;
 - variable selection methods sensible to select relevant markers, complexity of the “final” regression model depends on nominal significance level;
 - different variable selection strategies may result in different “final” regression models
 - 2b Trees:
 - uncritical application of trees can lead to large, instable and uninterpretable trees;
 - *P*-value correction and pre-specification of cutpoints for continuous variables
 - 2c Other approaches:
 - many methods available, not without problems.
3. *Formation of risk groups:*
 - different model building strategies may result in different risk groups;
 - avoid too small risk groups.
4. *Validation of results:*
 - overestimation of effects caused by data-dependent modelling;
 - validation of prognostic relevance of markers and models in independent validation study.
5. *Issues requiring more attention in future studies*
 - stability investigation of selected models;
 - combining variable selection with shrinkage;
 - differentiation between studies developing a prediction model from studies with main interest in one specific marker.

Figure 4 Summary of important issues for the analysis of single prognostic marker studies (Holländer and Sauerbrei, 2006).

2003). Simon (2001) commented that the prognostic literature ‘is probably cluttered with false-positive studies that would not have been submitted or published if the results had come out differently’. Such publication bias would thus lead to the literature being biased towards over-estimating the prognostic importance of markers.

Rifai *et al* (2008) believe that it is time to take action against reporting biases in prognostic studies. To encourage clear and transparent reporting of prognostic marker studies, some recent articles have provided reporting guidelines. The REMARK guidelines (McShane *et al*, 2005b) consider the whole study process and suggest the key information that needs reporting, from the pre-defined hypotheses and patients included, to the statistical analysis methods used and the results identified, and to the study limitations and implications for clinical practice. Journal editors are encouraged to enforce adherence to these guidelines before accepting a prognostic marker article. In addition, Burton and Altman (2004) consider how to report prognostic studies when there are missing covariate data, whereas Riley *et al* (2003) suggest how effect estimates and summary results should be reported (Figure 5), with a recommendation to also provide IPD to facilitate meta-analysis.

SYSTEMATIC REVIEWS AND META-ANALYSIS

Current difficulties for evidence synthesis

An evidence-based approach to prognostic markers is clearly needed. It is usually difficult to ascertain the benefit of a marker from a single published study, which may be overoptimistic owing

to small sample size and selective reporting, and a clear view is only likely to emerge from looking across multiple studies. However, clinicians do not have the time to review such a plethora of marker studies, and ideally need the most suitable prognostic markers to be identified for them, as exemplified by the American Society of Clinical Oncology in breast cancer (Harris *et al*, 2007). To aid this process, a systematic review is commonly used, which is an approach for identifying, evaluating and summarising an evidence-base. The term ‘systematic’ comes from the fact that the review process is performed using systematic and explicit methods, so that the review should be transparent and reproducible. If appropriate, a meta-analysis can be carried out at the end of a systematic review, which is a statistical approach that suitably combines the quantitative evidence from all the available studies (or from the subset of better-quality studies) to produce overall results for practice (Sutton *et al*, 2000).

Traditionally, meta-analysis methods use data extracted from published reports, but even for a summary assessment of a treatment effect from randomised trials, this approach has severe limitations (Piedbois and Buyse, 2004). Indeed, systematic reviews and meta-analyses of published prognostic studies generally highlight a confusing picture and are usually limited in their conclusions. As individual studies of prognostic markers are often poorly designed and poorly reported, the available evidence for synthesis is seriously limited. Also, multiple studies of a particular marker typically vary in important ways, such as the assays used, inclusion criteria for patients, types of treatment, cutpoint level, clinical outcomes, and statistical analysis. This methodological variation introduces heterogeneity between studies, and compounds the problem of summarising a marker’s prognostic importance and identifying how to implement it in practice. For

Essential to present:

- 1) The *hazard ratio* and its *confidence interval*, or the \log_e (hazard ratio) and its variance. Markers that have a continuous function should be modelled as a continuous variable using appropriate methods. If there is a justifiable reason for using a cutpoint for a continuous marker it should be specified at the start of the study and clearly reported.
- 2) The *number of patients* and *number of events in total*. For binary markers (and continuous markers if a cutpoint is used) also report the numbers within each group.
- 3) Both *unadjusted* and *adjusted* results for each marker. For adjusted results, clearly state what variables have been adjusted for. Ideally, a consistency in the set of adjustment factors used across studies should be sought through collaborative groups working toward prospectively planned pooled analyses. Otherwise, (i) always present results adjusted for age and stage of disease, and (ii) consider using the same set of adjustment factors as in important earlier studies.
- 4) *Individual patient data* in the paper or on the Internet. Alternatively, clearly indicate within the paper how to contact the group for IPD. Information about available data on markers that were not analysed should be included. Subject to any restrictions imposed by data protection laws and guidelines, include the following:
 - exact initial marker level and how marker was measured;
 - time of disease recurrence (if appropriate);
 - follow-up time;
 - final disease status;
 - levels of other existing prognostic markers of recognised and accepted importance for current clinical practice;
 - patient subgroup information, e.g., age, stage of disease, type of treatment received;
 - details of inclusion/exclusion criteria would also be beneficial.

Highly desirable to present:

- 5) *Exact P-values*. Reporting of results as 'significant' or 'not significant' is insufficient. Very small *P-values* can be given as $P < X$ (e.g., $P < 0.0001$), but in this case the exact χ^2 statistic is also needed.
- 6) Survival curves showing the difference in survival over time between the groups, with clear *step* and *censoring points*; also the initial numbers in each group, and the number of events and remaining numbers at various time points during follow-up are needed.
- 7) % *Survival at n years* with a confidence interval using Kaplan–Meier or other methods that allow for censoring, together with the number of patients at risk at that time in each group.

Figure 5 Guidelines for reporting the results of a prognostic marker study (Riley *et al*, 2003).

example, Parker *et al* (2001) performed a systematic review in prostate cancer to establish whether age is a prognostic marker, but the incomplete and heterogeneous nature of the reports prohibited any quantitative overview. Other issues that commonly hinder meta-analysis are the use of assays that are non-standardised or lacking in reproducibility; inappropriate or misleading statistical analyses; and optimal cutpoints, publication bias, and within-study selective reporting, all of which mean that the available published results are unlikely to represent a true picture. An assessment of the published results for marker MYCN in neuroblastoma strongly indicated that publication bias was present (Riley *et al*, 2004b), and the meta-analysis result obtained was likely to be an overestimate of the true prognostic effect for this marker. Furthermore, a recent review of the tumour suppressor protein, TP53, in head and neck cancer provides compelling empirical evidence that selective reporting biases are a major impediment to a meaningful meta-analysis (Kyzas *et al*, 2005). These biases have serious implications not only for meta-analysis but also for interpretation of the cancer prognostic literature as a whole (McShane *et al*, 2005a).

It is particularly important for systematic reviews of prognostic marker studies to appraise the methodological quality of the studies identified (Hayden *et al*, 2006). Often the main benefit of a systematic review of published prognostic studies is to expose the problems within primary research, and thus highlight the appropriate direction of future research. For example, Schmitz-

Dräger *et al* (2000) reviewed 43 trials that considered p53 immunohistochemistry as a prognostic marker in bladder cancer, and concluded that: 'From this analysis it becomes evident that further retrospective investigations will not contribute to the solution of the problem and thus are obsolete. There is an obvious need for standardisation of the assay procedure and the assessment of the specimens as well as for the initiation of a prospective multi-centre trial to provide definite answers.' Individual prognostic studies are usually carried out independently and not in the context of facilitating a systematic review or a meta-analysis, either during the study process or when writing for publication. New studies should rather build specifically on the results of previous studies, forming a collective drive towards answering questions of real clinical importance. In particular, when a study assesses the potential of a new prognostic marker, it should do so in relation to those other markers identified earlier as important and currently used in clinical practice. For instance, the stage and grade of disease are used as prognostic indicators in most tumours, and therefore results for new markers need to be adjusted for these variables. However, in general, such adjusted results are not always reported and, even when they are, great inconsistency exists across studies regarding which factors are used for adjustment (Riley *et al*, 2003). This and other quality issues regarding the analysis of primary studies can be circumvented if IPD are available rather than only the published results, as now considered.

- Use, as far as possible, consistent inclusion and exclusion criteria across studies, and if appropriate re-instate patients into the analysis who were originally excluded.
- Use up-to-date follow-up information, which is potentially longer than that used in the study publication.
- Obtain estimates for those missing or poorly reported outcomes and summary statistics across studies; it may thus reduce the problem of selective within-study reporting.
- Obtain more direct estimates (e.g., of the hazard ratio) where earlier only indirect estimates were available.
- As far as possible, standardise the strategy of statistical analysis across studies
- Produce adjusted estimates where earlier only unadjusted estimates were available.
- Use a (small) consistent set of adjustment factors across studies.
- Use a consistent cutpoint across studies, or produce continuous marker results.
- Assess the benefits of using combinations of markers.
- Assess specific subgroups of patients across studies (e.g., pre-menopausal, stage 4 disease), and assess whether patient level characteristics (such as, age and treatment) are effect modifiers across studies.
- Identify those studies that contain the same or overlapping sets of patients.
- Assess model assumptions in each study, such as proportional hazards.
- Produce appropriate statistical validation of the models developed.

Figure 6 Summary of the potential benefits of having individual patient data (IPD) for a meta-analysis of prognostic marker studies.

The benefit of IPD

The availability of IPD from primary prognostic studies offers many advantages for meta-analysis (Figure 6). For example, IPD allows the adequate checking of the data and modelling assumptions; unpublished results and outcomes to be obtained; an extended follow-up of patients; standardised analysis of the data with proper and consistent handling of continuous variables (Holländer and Sauerbrei, 2006; Royston *et al*, 2006); and suitable validation of the models developed (Altman and Royston, 2000; Holländer and Sauerbrei, 2006). It may also increase the opportunity to evaluate combinations of markers, which may produce more specific and accurate prognostic assessments than the individual markers themselves. We thus strongly recommend the IPD meta-analysis approach, but recognise that IPD may not solve all the problems (Stewart and Tierney, 2002), such as poorly designed primary studies. It may also be costly and time-consuming to obtain, or only available from a proportion of studies (Riley *et al*, 2007a, 2008).

The feasibility of obtaining IPD from prognostic studies has recently been considered by Altman *et al* (2006), and more generally for survival studies by Ioannidis *et al* (2002). Altman *et al* concluded that the IPD approach is possible, although it can be a 'long, expensive, and rather laborious process'. To reduce the time and cost required, one option is to seek IPD from only a 'well-defined list' of studies, as successfully done by Look *et al* (2002) in breast cancer, although this approach may be criticised for potentially subjective decisions regarding what constitutes a 'well-defined list' and a 'large' or 'high-quality' study.

TOWARDS MULTI-CENTRE AND COLLABORATIVE RESEARCH

Prospectively planned pooled analyses

It is clear that collaboration across multiple disciplines and multiple centres is required to achieve the necessary progress outlined in the sections 'Primary studies of prognostic markers' and 'Systematic reviews and meta-analysis'. However, such a collaborative drive naturally points away from the retrospective pooling of data, and points toward prospectively planned pooled analyses and prospective multi-centre studies. Indeed, McShane *et al*

(2005a) state that 'More importantly, the necessity of large, definitive prospective studies or prospectively planned meta-analyses for tumour marker research must be recognised.' Research groups working together and communicating from the outset of their studies – as protocols and clinical objectives are being formulated – are the best way to achieve consistency in, for example, study design, markers assessed, method of measurement, treatments considered, outcomes of importance, and statistical analysis. A pooled analysis could then be planned prospectively to answer pre-specified clinical questions, with investigators committing in advance to the availability of their IPD at the end of their study.

Prospective multi-centre studies are the ideal, but may be time-consuming and costly. For prognostic markers measurable on stored material, a helpful compromise could be a protocol-based pre-planned pooled analysis of retrospective data from several centres. Such a project would start with a detailed protocol specifying inclusion and exclusion criteria for patients and treatments allowed. The availability of complete data for some 'basic' variables and a minimum amount of follow-up could be inclusion criteria. Compared with the effect of markers, the effect of most treatments is often relatively small. Therefore, the variation of treatment across several centres will not matter too much. The situation is more complicated if the treatment effect depends strongly on the marker (treatment–covariate interaction). However, that is rare. Such a situation requires a more thoughtful analysis, for example, a stratified analysis or separate subgroup investigations. Ideally, in all centres, the same method of the marker measurement has to be used, otherwise methods to transfer data to one scale are required, and other technical variations across studies are restricted as far as possible. Combining the estimated effects from the individual centres (studies) using meta-analysis methodology or by stratifying the analysis by centre (or treatment) gives the flavour of a pre-planned meta-analysis, a design used for some time in epidemiology (Blettner *et al*, 1999; Boffetta *et al*, 2004; Cardis *et al*, 2005). A combined analysis of data from several centres (studies), which is based on a detailed protocol before data collection is started, is an important design for future prognostic marker research. The required sample size can be reached by adding more centres, and the time frame for such a study can be 'relatively' short as data from patients with long-term follow-up can be used. As thousands of patients are required for a reliable assessment of the importance

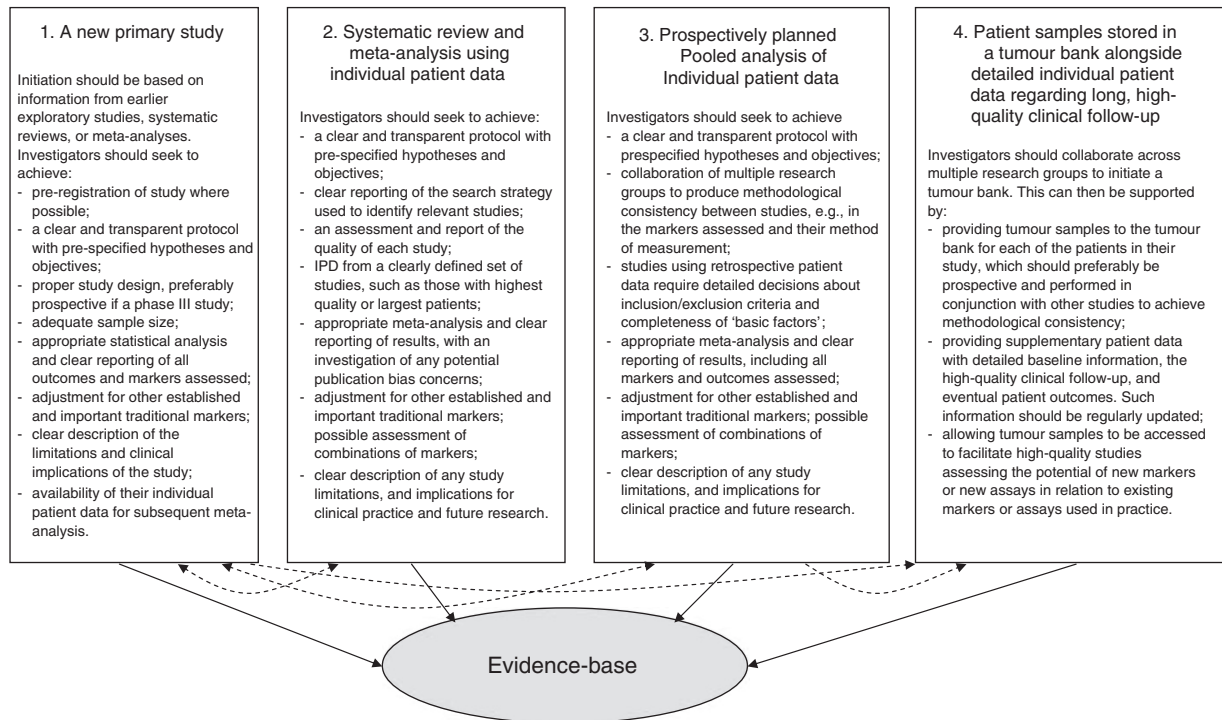


Figure 7 Pathways to high-quality evidence regarding the prognostic ability of a marker, following publication of initial hypothesis generating studies.

of a marker, it is necessary that large centres be willing to cooperate in such a project. Such an approach has many scientific advantages to a situation where large (and small) centres conduct studies on their own with hundreds of patients, but with variations in inclusion criteria, such as measurement techniques, analysis strategies, and so on. As obvious from the past, such unstructured type of research will not lead to scientific answers urgently needed. McShane *et al* (2005a) suggest 'cultural changes will be required', and the necessity of this type of collaboration has been recognised before by clinical trialists and by the epidemiologic community.

Tissue banks

Another promising possibility is the use of samples from a tissue bank, especially one that is established to facilitate prognostic research (Burke and Henson, 1998; Schilsky *et al*, 2002). Tumour samples and corresponding data may come from patients treated within a prospective (randomised) trial, or they may come from those not taking part in a trial. In the earlier case, storing may be connected with a more specific project for a specified population; in the latter case, patients may belong to a very heterogeneous population with regard to characteristics, treatment, and follow-up. Studies using the resource of a tissue bank would benefit from standardised collection and storage of samples, as well as from good quality baseline and follow-up data. If new assays need to be performed, they can be carried out using standardised laboratory methods. After the first few years, studies using such samples would have access to adequate clinical follow-up, and so could be carried out quickly but without the disadvantages associated with retrospective studies.

Hayes *et al* (2008) concur that the exciting potential of prognostic markers highlights the 'importance of prospective collection, processing, and storage of biospecimens' to help identify markers that facilitate individualised treatment strategies. A highly commendable example of such an initiative is in the bladder cancer field, where Goebell *et al* (2004) are establishing a multi-institutional bladder cancer database and a virtual tumour bank as

a resource for participating institutions to evaluate the biological and prognostic significance of potential markers. More generally, the Confederation of Cancer Biobanks (CCB) are developing biobank resources for cancer research and promoting consistency in how tissue samples are prepared and preserved, thus to facilitate research investigations involving a larger number of patient samples. Arguments against tissue banks usually relate to confidentiality or to medico-legal concerns, but increasingly tissue banks are accompanied by strong legal and ethical requirements, and require patient consent. A discussion of the current ethical issues involving tissue banks is reported by Ravid (2008).

Future single and multi-centre prospective prognostic marker studies can also facilitate the initiation of tissue banks by providing their patient tissue samples and archiving their IPD of each patient's baseline covariates, clinical follow-up, and eventual outcomes. This would allow further investigations in the light of new information after the studies ended; for example, if subsequently a new marker is discovered, then the tissue samples could be assessed for this marker and comparison made to the existing markers. Data from such a marker can be most relevant to the investigations of treatment-covariate interactions in a (single) large randomised trial, but it can also be used in prospectively planned pooled analyses as noted above. For a recent example investigating several markers for an interaction with treatment in a lung cancer trial, see Olausson *et al* (2006) and Filipits *et al* (2007a, b).

DISCUSSION

Any insight into the future health of an individual patient is advantageous, and so prognostic markers can potentially play a vital part in clinical decision-making (Windeler, 2000). It is imperative, therefore, that researchers produce reliable and informative evidence regarding the prognostic markers available for practice. However, for whatever reason, it is clear that prognostic marker studies have been neglected in the drive to improve medical research. The responsibility for this lies with all

those involved in prognostic research, from those identifying and measuring markers in the laboratory, to those designing and implementing primary studies, to those analysing and reporting results, and to those reviewing and publishing studies. We must now work together across multiple disciplines, and move towards transparent, high-quality, clinically relevant research relating to prognostic markers. Higher-quality prognostic marker research is desperately needed and is attainable. To achieve such progress, collaborative and multi-disciplinary teams should be an essential part of future prognostic studies, including clinicians, biologists, and statisticians, among others. Research on treatment strategies has successfully involved international and multi-disciplinary collaborations for many years, and this is undoubtedly the way forward for prognostic marker research. Indeed, the best chance of answering important questions about prognosis is likely to come from 'a systematic, sustained programme of epidemiologic research, coordinated among cognate research groups' (Hemingway, 2006). To this end, we encourage researchers to join the newly registered Prognosis Methods Group within the Cochrane Collaboration (Riley *et al*, 2007b).

In this paper, we have encouraged progress towards higher-quality prognostic studies by explaining where current deficiencies lie, and by discussing some existing guidelines regarding the design, clinical relevance, analysis, and reporting of such studies. Hemingway (2007) recommends that: 'We need not only to develop quality standards of primary, secondary, and tertiary prognosis research but also to secure their implementation'. Cultural changes thus need to be embedded in the research community, with continued dissemination of best practice made within and across research groups, and at national and interna-

tional meetings. A pivotal role in ensuring good practice is held by the editors of and reviewers for clinical journals, who can enforce certain standards. For example, some journals require the prospective registration of randomised controlled trials (De Angelis *et al*, 2004), and this is an option worth exploring further for prognostic studies. Researchers also need to look beyond single prognostic studies and consider the bigger picture: the clinical utility of a marker will only be established through multiple high-quality studies conducted over a period of time. Prospective phase III studies, prospectively planned meta-analyses of prospective or retrospective data, the availability of IPD, and the initiation of tumour banks all support the evolution of prognostic marker evidence. In Figure 7 we summarise this process and indicate the different stages in which researchers can make a significant contribution to a marker's evidence-base. Research groups should strive for such opportunities, and make a long-standing commitment towards high-quality and collaborative studies in the future. Only then will we achieve the evidence-based use of markers in practice and ensure the most appropriate patient care, which should be the aim of all of us.

ACKNOWLEDGEMENTS

Richard Riley was funded as a Research Scientist in Evidence Synthesis by the UK Department of Health National Coordinating Centre for Research and Capacity Development while undertaking this work. We thank the two reviewers whose thoughtful comments have helped improve the paper.

REFERENCES

- Altman DG (2001) Systematic reviews of evaluations of prognostic variables. *BMJ* **323**: 224–228
- Altman DG, De Stavola BL, Love SB, Stepniowska KA (1995) Review of survival analyses published in cancer journals. *Br J Cancer* **72**: 511–518
- Altman DG, Lausen B, Sauerbrei W, Schumacher M (1994) Dangers of using 'optimal' cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* **86**: 829–835
- Altman DG, Lyman GH (1998) Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Res Treat* **52**: 289–303
- Altman DG, Royston P (2000) What do we mean by validating a prognostic model? *Stat Med* **19**: 453–473
- Altman DG, Trivella M, Pezzella F, Harris AL, Pastorino U (2006) Systematic review of multiple studies of prognosis: the feasibility of obtaining individual patient data. I. In *Advances in Statistical Methods for the Health Sciences*, Auget J-L, Balakrishnan N, Mesbah M, Molenberghs G (eds), pp 3–18. Birkhäuser: Boston
- Bennett DA (2003) Review of analytical methods for prospective cohort studies using time to event data: single studies and implications for meta-analysis. *Stat Methods Med Res* **12**: 297–319
- Blettner M, Sauerbrei W, Schlehofer B, Scheuchenpflug T, Friedenreich C (1999) Traditional reviews, meta-analyses and pooled analyses in epidemiology. *Int J Epidemiol* **28**: 1–9
- Boffetta P, Soutar A, Cherrie JW, Granath F, Andersen A, Anttila A, Blettner M, Gaborieau V, Klug SJ, Langard S, Luce D, Merletti F, Miller B, Mirabelli D, Pukkala E, Adami HO, Weiderpass E (2004) Mortality among workers employed in the titanium dioxide production industry in Europe. *Cancer Causes Control* **15**: 697–706
- Boracchi P, Biganzoli E (2003) Markers of prognosis and response to treatment: ready for clinical use in oncology? A biostatistician's viewpoint. *Int J Biol Markers* **18**: 65–69
- Brundage MD, Davies D, Mackillop WJ (2002) Prognostic factors in non-small cell lung cancer: a decade of progress. *Chest* **122**: 1037–1057
- Burke HB, Henson DE (1998) Specimen banks for cancer prognostic factor research. *Arch Pathol Lab Med* **122**: 871–874
- Burton A, Altman DG (2004) Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. *Br J Cancer* **91**: 4–8
- Cardis E, Vrijheid M, Blettner M, Gilbert E, Hakama M, Hill C, Howe G, Kaldor J, Muirhead CR, Schubauer-Berigan M, Yoshimura T, Bermann F, Cowper G, Fix J, Hacker C, Heinmiller B, Marshall M, Thierry-Chef I, Utterback D, Ahn YO, Amoroso E, Ashmore P, Auvinen A, Bae JM, Solano JB, Biau A, Combalot E, Deboodt P, Diez Sacristan A, Eklof M, Engels H, Engholm G, Gulis G, Habib R, Holan K, Hyvonen H, Kerekes A, Kurtinaitis J, Malke H, Martuzzi M, Mastauskas A, Monnet A, Moser M, Pearce MS, Richardson DB, Rodriguez-Artalejo F, Rogel A, Tardy H, Telle-Lamberton M, Turai I, Usel M, Veress K (2005) Risk of cancer after low doses of ionising radiation: retrospective cohort study in 15 countries. *BMJ* **331**: 77
- De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, Kotzin S, Laine C, Marusic A, Overbeke AJ, Schroeder TV, Sox HC, Van Der Weyden MB (2004) Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *Lancet* **364**: 911–912
- Dupuy A, Simon RM (2007) Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* **99**: 147–157
- Filipits M, Haddad V, Schmid K, Huynh A, Dunant A, Andre F, Brambilla E, Stahel R, Pignon JP, Soria JC, Popper HH, Le Chevalier T, Pirker R (2007a) Multidrug resistance proteins do not predict benefit of adjuvant chemotherapy in patients with completely resected non-small cell lung cancer: International Adjuvant Lung Cancer Trial Biologic Program. *Clin Cancer Res* **13**: 3892–3898
- Filipits M, Pirker R, Dunant A, Lantuejoul S, Schmid K, Huynh A, Haddad V, Andre F, Stahel R, Pignon JP, Soria JC, Popper HH, Le Chevalier T, Brambilla E (2007b) Cell cycle regulators and outcome of adjuvant cisplatin-based chemotherapy in completely resected non-small-cell lung cancer: the International Adjuvant Lung Cancer Trial Biologic Program. *J Clin Oncol* **25**: 2735–2740
- Galea MH, Blamey RW, Elston CE, Ellis IO (1992) The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res Treat* **22**: 207–219
- Goebell PJ, Groshen S, Schmitz-Drager BJ, Sylvester R, Kogevinas M, Malats N, Sauter G, Barton Grossman H, Waldman F, Cote RJ (2004) The International Bladder Cancer Bank: proposal for a new study concept. *Urol Oncol* **22**: 277–284

- Harrell FE (2001) *Regression Modeling Strategies, with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer: New York
- Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, Somerfield MR, Hayes DF, Bast Jr RC (2007) American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* **25**: 5287–5312
- Hayden JA, Cote P, Bombardier C (2006) Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* **144**: 427–437
- Hayden JA, Côté P, Steenstra IA, Bombardier C, QUIPS-LBP Working Group (2008) Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. *J Clin Epidemiol* **61**: 552–560
- Hayes DF, Stearns V, Rae J, Flockhart D (2008) A model citizen? Is tamoxifen more effective than aromatase inhibitors if we pick the right patients? *J Natl Cancer Inst* **100**: 610–613
- Hemingway H (2006) Prognosis research: why is Dr. Lydgate still waiting? *J Clin Epidemiol* **59**: 1229–1238
- Hemingway H (2007) Improving prognosis research: standards primary, secondary, and tertiary. *J Clin Epidemiol* **60**: 865–866
- Henderson R, Keiding N (2005) Individual survival time prediction using statistical models. *J Med Ethics* **31**: 703–706
- Hinestroza MC, Dickersin K, Klein P, Mayer M, Noss K, Slamon D, Sledge G, Visco FM (2007) Shaping the future of biomarker research in breast cancer to ensure clinical relevance. *Nat Rev Cancer* **7**: 309–315
- Holländer N, Sauerbrei W (2006) On statistical approaches for the multivariable analysis of prognostic factor studies. In *Advances in Statistical Methods for the Health Sciences*. Auget J-L, Balakrishnan N, Mesbah M, Molenberghs G (eds), pp 19–38. Birkhäuser: Boston
- Horton R, Smith R (1999) Time to register randomised trials. The case is now unanswerable. *BMJ* **319**: 865–866
- Ioannidis JP, Rosenberg PS, Goedert JJ, O'Brien TR (2002) Commentary: meta-analysis of individual participants' data in genetic epidemiology. *Am J Epidemiol* **156**: 204–210
- Kattan MW (2003) Judging new markers by their ability to improve predictive accuracy. *J Natl Cancer Inst* **95**: 634–635
- Katz EM, Kattan MW (2005) How to judge a tumor marker. *Nat Clin Pract Oncol* **2**: 482–483
- Kyzas PA, Denaxa-Kyza D, Ioannidis JP (2007a) Almost all articles on cancer prognostic markers report statistically significant results. *Eur J Cancer* **43**: 2559–2579
- Kyzas PA, Denaxa-Kyza D, Ioannidis JP (2007b) Quality of reporting of cancer prognostic marker studies: association with reported prognostic effect. *J Natl Cancer Inst* **99**: 236–243
- Kyzas PA, Loizou KT, Ioannidis JP (2005) Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst* **97**: 1043–1055
- Look MP, van Putten WL, Duffy MJ, Harbeck N, Christensen IJ, Thomssen C, Kates R, Spyrtos F, Ferno M, Eppenberger-Castori S, Sweep CG, Ulm K, Peyrat JP, Martin PM, Magdelenat H, Brunner N, Duggan C, Lisboa BW, Bendahl PO, Quillien V, Daver A, Ricolleau G, Meijer-van Gelder ME, Manders P, Fiets WE, Blankenstein MA, Broet P, Romain S, Daxenbichler G, Windbichler G, Cufer T, Borstnar S, Kueng W, Beex LV, Klijn JG, O'Higgins N, Eppenberger U, Janicke F, Schmitt M, Foekens JA (2002) Pooled analysis of prognostic impact of urokinase-type plasminogen activator and its inhibitor PAI-1 in 8377 breast cancer patients. *J Natl Cancer Inst* **94**: 116–128
- Martin B, Paesmans M, Berghmans T, Branle F, Ghisdal L, Masciaux C, Meert AP, Steels E, Vallot F, Verdebout JM, Lafitte JJ, Sculier JP (2003) Role of Bcl-2 as a prognostic factor for survival in lung cancer: a systematic review of the literature with meta-analysis. *Br J Cancer* **89**: 55–64
- McShane LM, Altman DG, Sauerbrei W (2005a) Identification of clinically useful cancer prognostic factors: what are we missing? *J Natl Cancer Inst* **97**: 1023–1025
- McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM (2005b) REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* **93**: 387–391
- Olausson KA, Dunant A, Fouret P, Brambilla E, Andre F, Haddad V, Taranchon E, Filipits M, Pirker R, Popper HH, Stahel R, Sabatier L, Pignon JP, Tursz T, Le Chevalier R, Soria JC (2006) DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. *N Engl J Med* **355**: 983–991
- Parker CC, Gospodarowicz M, Warde P (2001) Does age influence the behaviour of localized prostate cancer? *BJU Int* **87**: 629–637
- Pencina MJ, D'Agostino SRB, D'Agostino JRB, Vasan RS (2008) Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* **27**: 157–172
- Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, Yasui Y (2001) Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* **93**: 1054–1061
- Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P (2004) Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* **159**: 882–890
- Piedbois P, Buyse M (2004) Meta-analyses based on abstracted data: a step in the right direction, but only a first step. *J Clin Oncol* **22**: 3839–3841
- Ransohoff R (2007) How to improve reliability and efficiency of research about molecular markers: roles of phases, guidelines, and study design. *J Clin Epidemiol* **60**: 1205–1219
- Ravid R (2008) Standard Operating Procedures, ethical and legal regulations in BTB (Brain/Tissue/Bio) banking: what is still missing? *Cell Tissue Bank* **9**: 121–137
- Rifai N, Altman DG, Bossuyt PM (2008) Reporting bias in diagnostic and prognostic studies: time for action. *Clin Chem* **54**: 1101–1103
- Riley RD, Abrams KR, Lambert PC, Sutton AJ, Altman DG (2006a) Where next for evidence synthesis of prognostic marker studies? Improving the quality and reporting of primary studies to facilitate clinically relevant evidence-based results. In *Advances in Statistical Methods for the Health Sciences*. Auget J-L, Balakrishnan N, Mesbah M, Molenberghs G (eds), pp 39–58. Birkhäuser: Boston
- Riley RD, Abrams KR, Sutton AJ, Lambert PC, Jones DR, Heney D, Burchill SA (2003) Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. *Br J Cancer* **88**: 1191–1198
- Riley RD, Heney D, Jones DR, Sutton AJ, Lambert PC, Abrams KR, Young B, Wailoo AJ, Burchill SA (2004a) A systematic review of molecular and biological tumor markers in neuroblastoma. *Clin Cancer Res* **10**: 4–12
- Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, Boutrif F (2008) Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med* **27**: 1870–1893
- Riley RD, Look MP, Simmonds MC (2007a) Combining individual patient data and aggregate data in evidence synthesis: a systematic review identified current practice and possible methods. *J Clin Epidemiol* **60**: 431–439
- Riley RD, Ridley G, Williams K, Altman DG, Hayden J, de Vet HC (2007b) Prognosis research: toward evidence-based results and a Cochrane methods group. *J Clin Epidemiol* **60**: 863–865
- Riley RD, Sutton AJ, Abrams KR, Lambert PC (2004b) Sensitivity analyses allowed more appropriate and reliable meta-analysis conclusions for multiple outcomes when missing data was present. *J Clin Epidemiol* **57**: 911–924
- Rothstein HR, Sutton AJ, Borenstein M (eds) (2005) *Publication Bias in Meta-Analysis*. Wiley: Chichester
- Royston P, Altman DG, Sauerbrei W (2006) Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* **25**: 127–141
- Royston P, Sauerbrei W (2008) *Multivariable Model-Building - A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. Wiley: New York
- Sargent DJ, Conley BA, Allegra C, Collette L (2005) Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol* **23**: 2020–2027
- Sauerbrei W (1999) The use of resampling methods to simplify regression models in medical statistics. *Appl Stat* **48**: 313–329
- Sauerbrei W (2005) Prognostic factors – confusion caused by bad quality of design, analysis and reporting of many studies. In *Current Research in Head and Neck Cancer: Advances in Oto-Rhino-Laryngology*. Bier H (ed), Vol. 62, pp 184–200. Karger: Basel
- Sauerbrei W, Holländer N, Riley RD, Altman DG (2006) Evidence-based assessment and application of prognostic markers: the long way from single studies to meta-analysis. *Commun Stat* **35**: 1333–1342
- Sauerbrei W, Hübner K, Schmoor C, Schumacher M (1997) Validation of existing and development of new prognostic classification schemes in node negative breast cancer. German Breast Cancer Study Group. *Breast Cancer Res Treat* **42**: 149–163
- Sauerbrei W, Royston P, Binder H (2007) Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med* **26**: 5512–5528
- Schilsky RL, Dressler LM, Bucci D, Monovich L, Jewell S, Suster S, Caligiuri MA, Kantoff PW, Compton C (2002) Cooperative group tissue banks as

- research resources: the cancer and leukemia group B tissue repositories. *Clin Cancer Res* **8**: 943–948
- Schmitz-Dräger BJ, Goebell PJ, Ebert T, Fradet Y (2000) p53 immunohistochemistry as a prognostic marker in bladder cancer. Playground for urology scientists? *Eur Urol* **38**: 691–699
- Schumacher M, Holländer N, Schwarzer G, Sauerbrei W (2006) Prognostic factor studies. In *Handbook of Statistics in Clinical Oncology*. Crowley J, Ankerst DP (eds), 2nd edn, pp 289–333. Chapman & Hall/CRC Press: Boca Raton, FL
- Schwarzer G, Vach W, Schumacher M (2000) On the misuse of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat Med* **19**: 541–561
- Simes RJ (1986) Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* **4**: 1529–1541
- Simon R (2001) Evaluating prognostic factor studies. In *Prognostic Factors in Cancer*. Gospodarowicz MK (ed), pp 49–56. New York: Wiley-Liss
- Simon R, Altman DG (1994) Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* **69**: 979–985
- Sterne JA, Egger M, Smith GD (2001) Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. *BMJ* **323**: 101–105
- Stewart LA, Tierney JF (2002) To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof* **25**: 76–97
- Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F (2000) *Methods for Meta-Analysis in Medical Research*. Wiley: London
- Tinker AV, Boussioutas A, Bowtell DD (2006) The challenges of gene expression microarrays for the study of human cancer. *Cancer Cell* **9**: 333–339
- Tudur-Smith C, Williamson PR, Marson AG (2005) Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Stat Med* **24**: 1307–1319
- Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R (2008) Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* **358**: 252–260
- Van Meerbeeck JP (1994) Prognostic factors in malignant mesothelioma: where do we go from here? *Eur Respir J* **7**: 1029–1031
- Vickers AJ, Elkin EB (2006) Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* **26**: 565–574
- Windeler J (2000) Prognosis – what does the clinician associate with this notion? *Stat Med* **19**: 425–430
- Wyatt J, Altman DG (1995) Commentary: Prognostic models: clinically useful or quickly forgotten? *BMJ* **311**: 1539–1541