British Journal of Cancer (2010) 102, 173 – 180

www.bjcancer.com

Reporting of prognostic studies of tumour markers: a review of published articles in relation to REMARK guidelines

S Mallett*[,1], A Timmer[2,3], W Sauerbrei[4] and DG Altman[1]

[1]Centre for Statistics in Medicine, University of Oxford, Oxford, UK; [2]Institute of Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, München, Germany; [3]German Cochrane Center, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg, Germany; [4]Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg, Germany

BACKGROUND: Poor reporting compromises the reliability and clinical value of prognostic tumour marker studies. We review articles to assess the reporting of patients and events using REMARK guidelines, at the time of guideline publication.

METHODS: We sampled 50 prognostic tumour marker studies from higher impact cancer journals between 2006 and 2007. The inclusion criteria were cancer; focus on single biological tumour marker; survival analysis; multivariable analysis; and not gene array or proteomic data. Articles were assessed for the REMARK profile and other REMARK guideline items. We propose a reporting aid, the REMARK profile, motivated by the CONSORT flowchart.

RESULTS: In 50 studies assessed for the REMARK profile, the number of eligible patients (56% of articles), excluded patients (54%) and patients in analyses (98%) was reported. Only 50% of articles reported the number of outcome events. In multivariable analyses, 54% and 30% of articles reported patient and event numbers for all variables. Of the studies, 66% used archival samples, indicating a potentially biased patient selection. Only 36% of studies reported clearly defined outcomes.

CONCLUSIONS: Good reporting is critical for the interpretability and clinical applicability of prognostic studies. Current reporting of key information, such as the number of outcome events in all patients and subgroups, is poor. Use of the REMARK profile would greatly improve reporting and enhance prognostic research.

British Journal of Cancer (2010) **102,** 173 – 180. doi:10.1038/sj.bjc.6605462 www.bjcancer.com
Published online 8 December 2009

Keywords: prognostic; REMARK; survival analysis; tumour marker; reporting guideline

Every year, thousands of articles are published on prognostic tumour markers, often with contradictory results for the same marker and disease. Many studies are so poorly reported that they lack the key information that readers need to evaluate their reliability and clinical applicability. It is of concern that health-care professionals may direct patient treatment on the basis of poorly reported studies. Systematic reviews of prognostic markers are often unable to include data from the majority of studies because of poor reporting (Riley *et al*, 2003, 2009; Malats *et al*, 2005; de Azambuja *et al*, 2007).

REMARK reporting guidelines (Table 1) were developed to encourage transparent and complete reporting in prognostic studies evaluating a single tumour marker on the basis of a recommendation from the National Cancer Institute and European Organisation for Research and Treatment of Cancer (McShane *et al*, 2005b). Reporting guidelines, such as the CONSORT guidelines for reporting randomised controlled trials (RCTs) (Moher *et al*, 2001), are important tools used by authors, journal editors and peer reviewers.

The variability in methods and conflicting results seen across prognostic studies makes reporting of key study details critical (Riley *et al*, 2003; Malats *et al*, 2005; Sauerbrei, 2005). Variability of study results can be partly attributed to patient spectrum/treatment, specimen characteristics, assay methods, study design and statistical analysis. The REMARK guidelines make specific reporting recommendations for each of these areas.

In addition, variation in prognostic studies is also caused by underpowered studies (Bentzen, 2001), multiplicity in analysis (Faraggi and Kramar, 2000), different cut points for markers (Sauerbrei, 2005), missing data (Burton and Altman, 2004), selective outcome reporting (Kyzas *et al*, 2005b) and publication bias (Kyzas *et al*, 2007a). These aspects of study design and statistical analysis might not be improved by better reporting, but might be more easily identified.

Systematic reviews in breast cancer (Altman, 2007), neuroblastoma (Riley *et al*, 2003), prostate cancer (Sutcliffe *et al*, 2009) and bladder cancer (Malats *et al*, 2005) found that poor reporting was common and limited the clinical applicability of studies. Often, only a minority of studies report sufficient information that can be considered for inclusion in a meta-analysis. In a systematic review of neuroblastoma, only 57 of 575 prognostic studies reported estimates for the hazard ratio or $\log_e$(hazard ratio) (Riley *et al*, 2003). Kyzas *et al* (2005a) found that evidence of

*Correspondence: Dr S Mallett, Centre for Statistics in Medicine, University of Oxford, Linton Rd, Oxford, OX2 6UD, UK; E-mail: susan.mallett@csm.ox.ac.uk

The study design was previously reported in the conference abstract. Mallett S, Altman DG, Timmer A, Sauerbrei W. REMARK profile study: reporting of participant flow and events in prognostic studies of tumour markers. Poster P12. XV Cochrane Colloquium, Sao Paulo, Brazil, 2007. Received 21 August 2009; revised 15 October 2009; accepted 19 October 2009; published online 8 December 2009

selective reporting bias is another serious issue. In an accompanying editorial, McShane *et al* (2005a) discussed the process of identifying clinically useful cancer prognostic markers and stressed that 'more complete and transparent reporting of marker studies would make it easier to distinguish carefully designed and analysed studies from haphazardly designed and over-analysed studies'.

Poor reporting of patient flow in studies and difficulties in getting an overview of all analyses performed have led to the development of a REMARK profile, inspired by the CONSORT flow diagram (Moher *et al*, 2001), to summarise key information on patient flow, study characteristics and statistical analyses. Table 2 shows a REMARK profile for an illustrative example. The REMARK profile will be published as a part of the forthcoming REMARK guideline explanatory document (personal communication REMARK guidelines group). The flow of patients through various analyses can be complicated in prognostic studies, particularly when there are missing data, multiple analyses or subgroup analyses. It is often difficult to identify all analyses. Results of several subgroup analyses are sometimes mentioned only very briefly, if at all.

In this article, we have conducted a review of the reporting of items from the REMARK guidelines in prognostic studies of tumour markers published in higher impact journals from January 2006 to April 2007. As the earliest publication of the REMARK guidelines was in August 2005, it was very unlikely that the guidelines were known to the authors when submitting the first version of their papers. Indeed, none of the papers referenced REMARK. Therefore, our study summarises the reporting of prognostic marker studies in the pre-REMARK era. In a following study, we intend to assess publications from 2009 using identical criteria. We assessed the REMARK reporting guideline items for patient characteristics, study design, data analysis and presentation of results, including the items in the REMARK profile.

MATERIALS AND METHODS

Literature search

A systematic hand search of five higher impact cancer journals (*Cancer*, *Cancer Research*, *International Journal of Cancer*, *Journal of Clinical Oncology*) that publish relevant articles was completed for 2006, and that for *Clinical Cancer Research* continued up to April 2007. The first 10 eligible articles for each journal were selected. We also planned to include the *European Journal of Cancer*, but found only four articles in 2006 that met our inclusion criteria and therefore we substituted *Cancer Research* instead.

Inclusion criteria

Articles were included if they examined the impact of a prognostic biological marker on at least one of overall survival (OS), disease-free survival (DFS), disease-free progression or recurrence in cancer patients, focused on one prognostic marker, but performed multivariable analysis with one or more additional variables.

Because of very different design and analysis issues, microarray, gene profiling and proteomics articles were excluded. The REMARK guidelines were developed for prognostic studies that evaluate a single tumour marker and are not designed for these study designs.

Biological markers included, for example, laboratory measurements, immunohistochemistry and DNA/RNA measurements, but did not include variables such as weight, BMI, angiogenesis measured by ultrasound or clinical tests such as reflex and lymph drainage pattern by scan.

Prognostic studies evaluating a single tumour marker were included irrespective of whether patients' data were from planned prospective trials, from clinical registries or other sources. We also did not have any limitation on the sample size of the study.

Table 1 Reporting recommendations for tumour marker prognostic studies (REMARK)

INTRODUCTION	
1.	State the marker examined, the study objectives, and any pre-specified hypotheses.
MATERIALS AND METHODS	
Patients	
2.[a]	Describe the characteristics (e.g., disease stage or comorbidities) of the study patients, including their source and inclusion and exclusion criteria.
3.[a]	Describe treatments received and how chosen (e.g., randomised or rule-based).
Specimen characteristics	
4.	Describe type of biological material used (including control samples) and methods of preservation and storage.
Assay methods	
5.	Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study end point.
Study design	
6.[a]	State the method of case selection, including whether prospective or retrospective and whether stratification or matching (e.g., by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.
7.[a]	Precisely define all clinical end points examined.
8.[a]	List all candidate variables initially examined or considered for inclusion in models.
9.	Give rationale for sample size; if the study was designed to detect a specified effect size, give the target power and effect size.
Statistical analysis methods	
10.[a]	Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.
11.	Clarify how marker values were handled in the analyses; if relevant, describe methods used for cut point determination.
RESULTS	
Data	
12.[a]	Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined report the number of patients and the number of events.
13.	Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumour marker, including numbers of missing values.
Analysis and presentation	
14.[a]	Show the relation of the marker to standard prognostic variables.
15.[a]	Present univariate analyses showing the relation between the marker and outcome, with the estimated effect (e.g., hazard ratio and survival probability). Preferably provide similar analyses for all other variables being analysed. For the effect of a tumour marker on a time-to-event outcome, a Kaplan–Meier plot is recommended.
16.[a]	For key multivariable analyses, report estimated effects (e.g., hazard ratio) with confidence intervals for the marker and, at least for the final model, all other variables in the model.
17.[a]	Among reported results, provide estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their statistical significance.
18.	If done, report results of further investigations, such as checking assumptions, sensitivity analyses and internal validation.
DISCUSSION	
19.	Interpret the results in the context of the pre-specified hypotheses and other relevant studies; include a discussion of limitations of the study.
20.	Discuss implications for future research and clinical value.

[a]Items assessed in this study. (produced permission of authors of McShane LM, Altman DG, Sauerbrei W *et al.* from *Br J Cancer* 2005; 93: 387–391).

Table 2 REMARK profile of patients, variables and statistical analyses (Study profile for Pfisterer *et al*, 1994). The REMARK profile is shown for illustrative purposes in an adaptable format. Additional rows can be included for each multivariable analysis, subgroup analysis or further outcome investigated

(a) Patients, treatment and variables

Study and marker	**Remarks**
Marker (If non-binary: how was marker analysed? continuous or categorical. If categorical, how are cutpoints determined?)	M = ploidy (diploid, aneuploid)
Further variables (variables collected, variables available for analysis, baseline variables, patient and tumour variables)	v1 = age, v2 = histological type, v3 = grade, v4 = residual tumour, v5 = stage, v6 = ascites[a], v7 = oestrogen[a], v8 = progesterone[a], v9 = CA-125[a]

Patients	***n***	**Remarks**
Assessed for eligibility	257	*Disease*: advanced ovarian cancer, stage III and IV *Patient source*: Surgery 1982–1990, University Medical Center Freiburg *Sample source*: archived specimens available
Excluded	73	General exclusion criteria[b], non-standard therapy[b], CV > 7%[b]
Included	184	Previously untreated. *Treatment*: all had platinum-based chemotherapy after surgery
With outcome events	139	Overall survival: death from any cause

(b) Statistical analyses of survival outcomes

Analysis	**Patients**	**Events**	**Variables considered**	**Results/remarks**
A1: Univariable (Provide for all variables. Give numbers as range if variables have different numbers of missing values)	184	139	M,v1 to v5	Tab 2, Fig1
A2: Multivariable	174	133	M,v1,v3 to v5	Tab 3 (v2 omitted because of many missing data; backward selection, see text)
A3: Effect for ploidy adjusted for v4	184	139	M, 4	Fig 2 (based on the result of A2)
A4: Interaction ploidy and stage	175	133	M, v1, v2, v4, v5	See text
A5: Ploidy in stage subgroups				
v5 = III	128	88	M	Fig 3
v5 = IV	56	51	M	Fig 4

[a]Not considered for survival outcome as these factors are not considered as 'standard' factors and/or number of missing values are relatively large. [b]Values not given in the paper.

One reader (SM) selected articles for inclusion. Queries on article inclusions were referred to second readers (AT, WS, DA).

Validity assessment and data abstraction

We assessed 50 articles in random order using a pre-piloted data extraction form of 52 items based on the REMARK guidelines (McShane *et al*, 2005b) covering study, patient and article characteristics; definition of study outcomes; and univariable and multivariable analyses. The form is available on request. Two reviewers (SM and AT) completed duplicate data extraction from 15 articles (two articles were subsequently excluded as the journal was excluded) with reference to a third reader where necessary. Differences were mostly due to difficulty in locating information in articles or use of form. Modifications to the data extraction form improved reliability. Agreement between readers was 89% for the last five articles, with disagreements due to difficulties in finding information or ambiguity in articles. As pre-specified for high agreement, a single reader (SM) completed the remaining articles, referring queries to another reader. Reporting of treatment was problematic and two readers reviewed the text from all articles.

Items in the REMARK profile were assessed for one outcome per article. Table 2 provides a suggested reporting format for the REMARK profile, with the REMARK profile items illustrated using data from an example article (Pfisterer *et al*, 1994). If there was more than one disease or multivariable outcome in the article, the first reported in the title, abstract or text was selected. Within each article, the same outcome was assessed for univariable and multivariable analyses.

Table 3 Reporting of patient and event numbers

REMARK profile item	**Number of articles reporting % (*n* = 50)**
Number of patients overall	
Assessed for eligibility	56 (28)
Excluded	54 (27)
Number available for analysis	
Patients	98 (49)
Events	50 (25)
Numbers in univariable analysis[a]	
Patients	54 (27)
Events	21 (11)
Numbers in multivariable analysis[b]	
Patients	54 (27)
Events	30 (15)
Median number of items reported[c]	4

[a]Only one univariable outcome per article, but univariable analyses for all variables are assessed for this outcome. Univariable outcome is same as outcome used as for multivariable analysis. [b]Only one multivariable analysis assessed. [c]Median of these eight items.

Overview of reporting the number of patients and events

A score was developed to allow a visual representation of reporting eight key items per article from the REMARK profile (Table 3). This score was used solely for descriptive purposes in this study, recognising the fact that not all information has equal importance.

The reporting of each item was assigned a score of 1 if clearly reported, and zero if unclear or not reported. Only one outcome was assessed per article for univariable and multivariable analyses for simplicity of presentation. An analysis based on multiple outcomes per article did not materially change the results. Data were plotted in Stata 10.0 (StataCorp, College Station, TX, USA).

RESULTS

Figure 1 shows a flow diagram of included articles. A full list is presented in a Supplementary Table. The cancer sites included are breast (10 articles); colorectal (6); urological (6); skin (6); ovary (5); head and neck (5); brain and nervous system (4); upper GI and pancreas (4); haematological malignancies (2); endometrium (1); and bone and soft tissue (1).

A total of 44 different primary prognostic markers were investigated in 50 articles, with six markers each studied in two articles. Of these six markers, four were presented for different cancer sites and two for the same site; one marker was reported in two articles by the same authors at the same site. Of the articles, 54% (27) used immunohistochemistry to identify the primary marker, 22% (11) used PCR, 12% (6) used ELISA, 4%(2) used microscopic scoring but not immunohistochemistry and 6% (3) used other methods. In one article, the method was unclear.

Tables 3, 4 and 5 show the findings across the areas assessed.

Reporting of patients and events

The numbers of patients and events are key items in the REMARK profile that provide an overview of patient flow in a study. The number of events in analyses is the 'effective' sample size and dictates the variability of study estimates.

We assessed key items from the REMARK profile that relate to the number of patients and events based on a single outcome per article (Table 3). Only 50% (25) of articles reported the number of events included in analyses, with no difference observed between OS and DFS outcomes. The number of excluded patients was reported in seven studies, could be calculated in 20 studies, but was unavailable in 23 studies because of a lack of reporting of eligible patients (22) or included patients (1). In 22 of 27 studies for which data were available, the number of patients included in analyses was fewer than the number of eligible patients.

If data are missing, a full reporting of univariable analysis with survival outcomes requires both patient and event numbers for all variables analysed, not just the biomarker of interest. Analyses should indicate the amount of missing data for covariates.

We assessed the reporting of patient and event numbers for all variables in univariable analyses. For the primary study marker, 98% (49) of articles reported patient numbers and 42% (21) reported event numbers. Only 56% of articles reported the explicit univariable analysis of variables other than the primary marker. For all articles, regardless of their explicit reporting of the univariable analysis of all variables, 55% (27) of articles reported patient numbers for variables available for univariable analyses but only 11 reported event numbers (Table 3).

For multivariable analyses, 54% (27) and 30% (15) of articles reported patient and event numbers, respectively, including 12 studies without missing data in which all patients available for analysis were included in the multivariable analysis so that numbers could be inferred even if not specifically reported for the multivariable analysis.

Figure 2 shows a graphical overview of the items reported per article, using star plots. Each article is represented by a star, with the eight spokes corresponding to the patient and event items from Table 3. Only articles 1–7 reported all eight items. Articles reporting fewer items on patient and event numbers from the REMARK profile tended to report patient numbers rather than event numbers. Overall, half (50%) of all articles did not provide event numbers for any reported outcomes (spokes at 8, 9, and 10 o'clock).

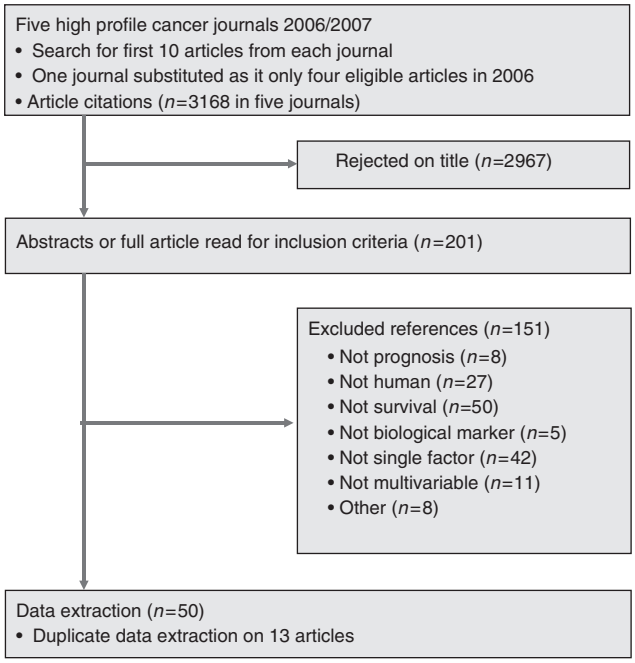

Figure 1 Flowchart of included articles.

Table 4 Patient, study and outcome reporting ($n = 50$)

Topic	Items reported	% (n) articles
Patients	Source of patients (clinical setting/clinical trial)	82 (41)
	Age[a]	60 (30)
	Stage or grade of patients	92 (46)
	Selection of patients?	
	Apparently unselected	8 (4)
	Selected, some criteria given	56 (28)
	Unclear	36 (18)
Study	Start date of patient recruitment[b]	74 (37)
	Finish date of patient recruitment	74 (37)
	End of follow-up date?	18 (9)
	Median follow-up for patients	58 (29)
	Completeness of followup	26 (13)
Outcome	Outcomes examined	
	OS and DFS	46 (23)
	OS only	46 (23)
	DFS only	8 (4)
	Definition of multivariable outcomes	
	Multivariable outcome clearly defined[c]	36 (18)
	OS ($n = 29$)	
	Explicitly any death	2 (1)
	Cancer death only	20 (10)
	Type of death unclear	36 (18)
	DFS ($n = 19$)	
	DFS including deaths[c]	14 (7)
	DFS not including deaths	0 (0)
	DFS but unclear if includes deaths	24 (12)
	Multivariable outcome unclear	6 (3)

[a]Mean, or median age plus age range. [b]Ten articles reported dates spanning more than 10 years. [c]One multivariable outcome assessed per article. This includes two articles that did not define type of death for DFS.

Table 5 Analysis methods and estimates ($n = 50$ articles)

Topic	Items reported	% (*n*) articles
Analysis method	Cox only	96 (48)
	Cox and logistic regression	2 (1)
	Not reported	2 (1)
	Assumptions of proportional hazards examined?	8 (4)
	Is the relationship of the primary marker with the standard prognostic variables shown?[a]	80 (40)
Univariable analysis	*Primary marker*	
	Effect estimate (e.g. HR)[b]	58 (29)
	CI for effect estimate	42 (21)
	P-value for the marker	96 (48)
	KM graph by the primary marker	98 (49)
	Other variables	
	Explicit other univariable analyses	56 (28)
	Effect estimates for markers (e.g. HR)[b]	38 (19)
	CI for effect estimates	24 (12)
Multivariable analysis	More than one multivariable analysis reported	60 (30)
	Effect estimate for primary marker	84 (42)
	Effect estimate for other variables in the final model	
	All	66 (33)
	Some	10 (5)
	CI for effect estimates[c]	84 (42)
	P-value for marker	
	All variables	72 (36)
	Some variables	22 (11)
	KM graph for adjusted effect of the marker	0 (0)

[a]Yes if age, stage or grade is considered. [b]10 articles reported % 5-year OS or DFS as effect estimate. [c]This includes six articles that reported 95% CI for only one variable and five that reported for the primary marker only.

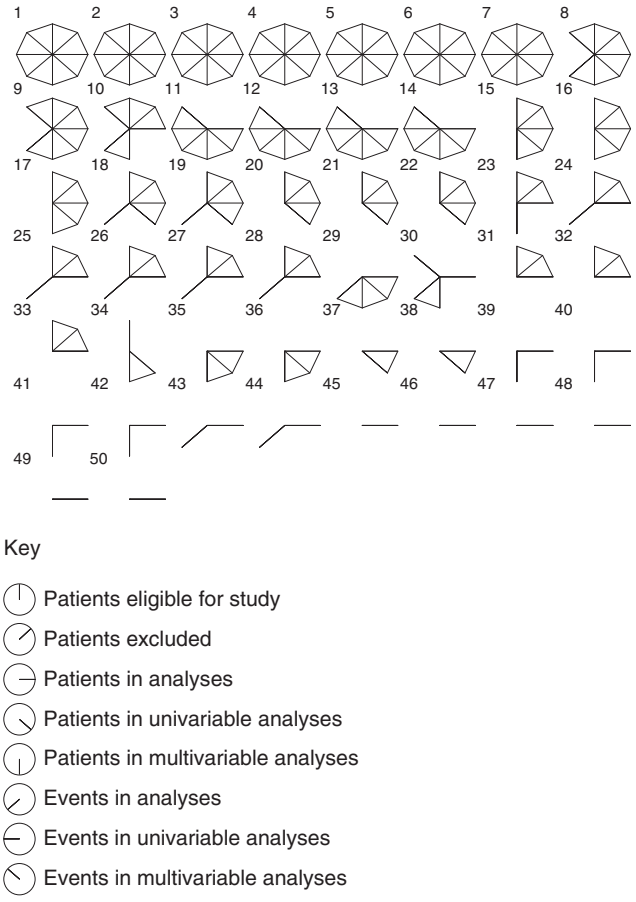

Figure 2 Reporting of patient and event numbers. Each article is represented by a star, with the eight spokes corresponding to the patient and event items detailed in the key below and in Table 1. Spokes are ordered in such a manner that reporting of patient event numbers is displayed at 8, 9 and 10 o'clock. Articles are ordered by the total number of items reported.

Reporting of patient characteristics

REMARK items 2, 3 and 6 recommend reporting of patient characteristics, treatments received and methods of patient selection. These are key to understanding the transferability and applicability of study results and the potential for study bias (Hayden *et al*, 2006).

The source of patients was reported in 82% (41) of articles (Table 4). Key patient characteristics in prognostic cancer research usually include age and disease severity (stage or grade of cancer). We defined clear reporting of age as including mean or median age plus age range, reported in only 60% (30) of articles. Of the articles, 92% (46) reported disease severity.

Heterogeneity in diagnosis and treatment of patients is common within prognostic studies because of the use of existing clinical patient databases (Pfisterer *et al*, 1994; Gasparini, 1998; Sauerbrei, 2005). We attempted to assess the reporting of patient treatment (item 3, Table 1). Most papers provided some information, but it was difficult to assess what constituted good reporting because of multiple treatments, different treatment time points relative to surgery, our lack of specialist knowledge of different cancers and the level of detail reported. We were unable to define a robust way to formally assess the reporting of treatment and how it was determined, across this range of cancer sites. Examples of details included treatments offered under local or national recommendations, but often not treatment uptake. When details were given, it was often not possible to judge how many patients received treatment, as seen in the following example 'The tumours were radically resected if possible and most patients with high-grade gliomas also received radiotherapy.' (Haapasalo *et al*, 2006). In all, 32% (16) of articles included treatment as a variable in univariable or multivariable analyses, or as stratification in the multivariable model.

Selection of patients from clinical populations was explicitly reported in 56% (28) of articles, with some selection criteria reported. In 36% (18) of articles, it was unclear whether patients were selected, as no criteria were reported; however, often selection was indicated implicitly as archival tumour bank samples were used. In 8% (4) of articles, patients were apparently unselected other than by time, as patients were enrolled from prospective trials with no reported exclusions (two studies), were consecutive patients (one study) or all of the consented patients (one study). The most frequently cited explicit or implicit selection criteria were availability of patient samples (43 articles), covariate data (12), clinical follow-up (7), disease/disease stage (5), treatment (3), level of primary marker (3) and comorbidity (2). Articles often cited several selection criteria. In 66% (33) of articles, samples were from an archive, whereas in 34% (17), samples were apparently collected for the specific study, including five articles in which samples were from phase II trials or RCTs.

Reporting of study characteristics

Item 6 of the REMARK guidelines recommends reporting key study dates. Clinical interpretation of a study depends on its time frame. Of the articles, 74% (37) reported both the start and

finish date of patient recruitment. The end of patient follow-up date was reported in only 18% (9) of studies (Table 4).

Loss to follow-up and completeness of study data are important aspects to assess the potential for study attrition bias (Schemper and Smith, 1996; Clark *et al*, 2002). Median follow-up was reported in 58% (29) studies, with a statement pertaining to completeness of follow-up in only 26% (13) studies.

Study outcomes and their definition

Item 7 of REMARK specifies the explicit definition of clinical end points, which is required for the interpretation of study results (Altman *et al*, 1995). With OS, it should be clear whether events refer to all deaths or only cancer deaths. For DFS, disease response or disease progression events need precise definition, as well as whether deaths are included as events, and which type of death.

Of the articles, 92% (46) examined OS; in half of them (23), this was the only outcome examined (Table 4). A total of 54% (27) of articles examined DFS, most of which (23) also included OS.

We assessed the definition of the outcome measure used in the first multivariable analysis in each article (see Materials and methods). Only 36% (18) of studies provided a clear definition of the outcome (Table 4). In articles using DFS as an outcome, seven articles included death as an event, but two of these did not clarify whether this event consists of all deaths that occurred or cancer deaths only.

Study variables

Item 8 of REMARK requires a list of all candidate variables initially examined or considered for inclusion in models. Over-optimistic models can result from multiplicity when significance testing is used on a large number of candidate variables, particularly with the small study sizes found in this sample (Altman, 2006).

The median number of variables investigated in the multivariable model was 8 (IQR 6–8, range 3–19). In eight studies, the number of variables was unclear; hence, we used the number of variables reported in tables, which is likely to be an underestimate. A total of 72% (36) of studies included age and 98% (49) included disease severity as study variables.

Size of studies

We extracted the size of studies in our sample of 50 papers. Of the articles, 98% (49) reported the overall number of patients, therefore our estimate of median size is reliable. However, only 50% (25) of articles reported the number of events for any outcome, therefore these estimates could be subject to reporting bias. We speculate that the number of events is smaller in articles not providing these numbers.

The median number of patients per study was 136 (IQR 77–234) range 43–889. The median number of events for OS was 72 (IQR 31–116, range 6–312, $n=23$ out of 46). The median number of events for DFS was 38 (IQR 23–71, range 17–280, $n=11$ out of 27).

Estimates of effect size and uncertainty

Items 15, 16 and 17 of REMARK (McShane *et al*, 2005b) recommend the reporting of estimated effects (e.g., hazard ratio and survival probability) with confidence intervals for all variables in univariable and multivariable analyses. This facilitates the inclusion of study results in systematic reviews. A Kaplan–Meier plot is recommended to present the association of a marker with time-to-event outcomes.

Estimates of the effect size for the primary marker were reported for univariable analyses in 58% (29) of articles, 42% with confidence intervals, whereas 96% reported *P*-values only (Table 5). In all, 98% (49) of articles presented Kaplan–Meier plots for the primary marker. Effect estimates and confidence intervals for variables other than the primary marker were reported less frequently.

In the multivariable analysis, effect estimates were more frequently reported, in 84 and 66% of studies for the primary marker and all markers, respectively.

Analysis methods

Item 10 (McShane *et al*, 2005b) includes recommendations to report statistical model methods and testing of assumptions. Item 14 requests the reporting of the relation of the marker to standard prognostic variables.

A total of 98% (49) of articles used the Cox proportional hazards method (Table 5). In all, 8% (4) of articles reported testing the assumption of proportional hazards and 80% of studies reported the relationship of the primary marker to at least one of age, stage or grade.

Examples of better reporting

We highlight three examples of better reporting, in which between 23 and 28 items of 39 items extracted in our study assessment were reported from the REMARK profile and from the items detailed in Tables 4 and 5 (Huang *et al*, 2006; Wadehra *et al*, 2006; Wang *et al*, 2006), with the proviso that in these studies outcome definitions are not explicitly described.

DISCUSSION

This study provides evidence of poor reporting in the current prognostic literature, and supports the potential value of using the REMARK checklist and REMARK profile to improve reporting. Poor reporting is a major obstacle to the applicability, transparency and evidence-based clinical use of current prognostic marker studies to direct patient treatment (Sauerbrei, 2005; Riley *et al*, 2006; Sauerbrei *et al*, 2006).

The REMARK guidelines were developed to improve reporting in prognostic studies of tumour markers by clearly indicating those events that are considered to be relevant and important enough to be reported so that poor reporting can be easily identified by non-specialist journal editors, peer reviewers, readers and authors.

REMARK profile

Our review assesses the reporting of items of the REMARK profile, a rapid overview of patients and events in a prognostic study designed to accompany the REMARK guidelines, similar to the CONSORT flowchart for RCTs. In addition, it gives an overview of the multiplicity of analyses conducted in each study. Table 2 shows an illustrative example from a study by Pfisterer *et al* (1994) in an adaptable format. Additional rows can be included for each multivariable analysis, subgroup analysis or for further outcomes investigated.

This research is new, as the REMARK profile assesses an overall picture of patients and events reported across studies as a whole. We found that a typical article only reported half of the REMARK profile items and these were often difficult to find. Half of the articles (25 out of 50) did not report the number of events for any analyses or outcomes. Studies in our sample had low sample sizes, with a median of 72 (IQR 31–116) events for OS and 38 (IQR 23–71) for DFS outcomes. Articles included a median of 136 patients (IQR 77–234).

Previous research has assessed parts of the REMARK profile, but has not addressed the flow of patients in a prognostic study as a whole. A review of prognostic studies on liver cirrhosis found that excluded patients were reported in one of 13 studies (Infante-Rivard *et al*, 1989). A review of survival analyses found that 93% of

articles reported the number of included patients, but only 45% of papers gave the number of events for each outcome (Altman *et al*, 1995). Studies of p53 were found to have a mean size of 86 patients, with 70% of studies enrolling 100 or fewer patients (Malats *et al*, 2005). Similarly, 83% of TP53 studies were of fewer than 100 patients (Kyzas *et al*, 2005b). Typically, prognostic studies are too small to be reliable, either to detect important prognostic factors or to provide substantial evidence for factors identified (Altman and Lyman, 1998; Faraggi and Kramar, 2000; Bentzen, 2001; Altman and Riley, 2005; Sauerbrei *et al*, 2006).

The source of patients was reported in 82% of articles, with 92% of articles indicating selection of patients from those populations. Of the studies in our sample, 66% indicated selection due to availability of patient specimens from hospital archives. Similarly, specimen or test availability was reported as selection in inclusion criteria in 40% of articles (Burton and Altman, 2004) Selection bias is common when archival tumour samples are used, as sample collection is a part of diagnosis and treatment, and is guided by disease severity (Hoppin *et al*, 2002). Determinants of selection and their implications for generalisability and interpretation of prognostic findings need to be reported.

Other REMARK items

Our study included several items from REMARK guidelines that are not in the REMARK profile. These have been investigated in previous research in other areas of prognosis, including systematic reviews of individual diseases or tumour markers, and have also found evidence or discuss poor reporting of follow-up time (Marx and Marx, 1997; Malats *et al*, 2005; Altman, 2007); loss of follow-up (Infante-Rivard *et al*, 1989; Burton and Altman, 2004; Altman, 2007); study dates (Altman, 2007); effect estimates and confidence intervals for univariable and multivariable analyses (Riley *et al*, 2003; Scholten-Peeters *et al*, 2003; Barth *et al*, 2004; Kuijpers *et al*, 2004; Malats *et al*, 2005); definitions of outcomes of overall survival and disease-free survival (Altman *et al*, 1995; Malats *et al*, 2005; Hudis *et al*, 2007; Kyzas *et al*, 2007b); and heterogeneity of patient and patient treatments (Pfisterer *et al*, 1994; Gasparini, 1998; Sauerbrei, 2005).

We assessed the reporting of items in the REMARK guidelines for prognostic studies with a focus on a single tumour marker. A complementary study from Kyzas *et al* (2007b) assessed reporting on the study design and assay methods of the REMARK guidelines. Similar findings of poor reporting has been found in other prognostic articles and other medical areas (Scholten-Peeters *et al*, 2003; Barth *et al*, 2004; Kuijpers *et al*, 2004). We studied higher impact articles, but poor reporting has been found across other journals and study types (Altman, 2007; Kyzas *et al*, 2007b). Previous to the widespread availability of journal webspace for additional information, it could have been argued that some aspects of poor reporting were necessitated by word restrictions. However, key items such as number of events would require only a few words. The reliability of our assessment was good, as we used duplicate data extraction until data extraction had a high level of agreement between two independent readers.

Good reporting cannot make up for poor study design; however, it can facilitate the identification of poor studies. There is plenty of evidence that poor study design and analysis are widespread in prognostic studies (Altman and Lyman, 1998; Altman and Riley, 2005; Sauerbrei, 2005). In addition, serious concerns about the high level of reporting bias has been raised (Kyzas *et al*, 2007a).

We found 44 different prognostic markers in 50 articles. Despite the limited number of prognostic factors used in standard clinical practice, the prognostic literature is replete with prognostic markers. Systematic reviews even within single diseases, neuroblastoma and non-small lung cancer, have identified 130 and 169 different prognostic markers, respectively (Brundage *et al*, 2002; Riley *et al*, 2003).

Assessing 10 articles in each of five journals does not allow a reliable comparison between the reporting quality of journals. There was some variation in reporting between the five journals, but more important is that all journals published articles with wide variations in the quality of reporting.

The wide variation and poor quality of reporting within these and other journal articles suggest that adherence to the REMARK guidelines would produce a framework to facilitate author, peer reviewer and editor consistency. We note that many journals, including three journals assessed in this study (*Cancer*, *Cancer Research* and *International Journal of Cancer*), do not mention or ask for adherence to REMARK guidelines for tumour marker studies in their instructions to authors (checked 6 October 2009).

As the REMARK guidelines were first published in June 2005, our study summarises the reporting of prognostic marker studies in the pre-REMARK era. In a following study, we intend to assess publications from 2009 using identical criteria.

In conclusion, we assessed the reporting of elements of the REMARK guidelines that relate to patients and events in published reports of tumour marker prognostic studies. Recently, the REMARK guidelines have been used in a systematic review of prognostic studies in stroke (Whiteley *et al*, 2009). Poor reporting of patient and event numbers had been anticipated, but we were dismayed that less than half of the articles reported event numbers, so readers had no information on an effective study sample size. We emphasise the importance of the REMARK profile as a standardised format for presenting key details of a tumour marker prognostic study. We hope that if journals required authors to include a REMARK study profile, those articles would have enhanced transparency and value for clinicians and patients.

ACKNOWLEDGEMENTS

In this study, SM and DGA were funded by Cancer Research UK and AT was funded by BMG (Federal Ministry of Health).

Conflict of interest
The authors declare no conflict of interest.

Supplementary Information accompanies the paper on British Journal of Cancer website (http://www.nature.com/bjc)

REFERENCES

Altman DG (2006) Studies investigating prognostic factors: Conduct and evaluation. In *Prognostic Factors in Cancer* Gospodarowicz MK, O'Sullivan B, Sobin LH (eds) 3rd edn, pp 39–54. John Wiley & Sons: New York

Altman DG (2007) Prognostic models: a methodological framework and review of models for breast cancer. In *Breast cancer. Translational therapeutic strategies* Lyman GH, Burstein HJ (eds) pp 11–25. Informa Healthcare: New York

Altman DG, De Stavola BL, Love SB, Stepniewska KA (1995) Review of survival analyses published in cancer journals. *Br J Cancer* **72:** 511–518

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, Riley RD (2005) Primer: an evidence-based approach to prognostic markers. *Nat Clin Pract Oncol* **2:** 466 – 472

Barth J, Schumacher M, Herrmann-Lingen C (2004) Depression as a risk factor for mortality in patients with coronary heart disease: a meta-analysis. *Psychosom Med* **66:** 802 – 813

Bentzen SM (2001) Prognostic factor studies in oncology: osteosarcoma as a clinical example. *Int J Radiat Oncol Biol Phys* **49:** 513 – 518

Brundage MD, Davies D, Mackillop WJ (2002) Prognostic factors in non-small cell lung cancer: a decade of progress. *Chest* **122:** 1037 – 1057

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

Clark TG, Altman DG, De Stavola BL (2002) Quantification of the completeness of follow-up. *Lancet* **359:** 1309 – 1310

de Azambuja E, Cardoso F, de Jr CG., Colozza M, Mano MS, Durbecq V, Sotiriou C, Larsimont D, Piccart-Gebhart MJ, Paesmans M (2007) Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *Br J Cancer* **96:** 1504 – 1513

Faraggi D, Kramar A (2000) Methodological issues associated with tumor marker development. *Biostatistical aspects. Urol Oncol* **5:** 211 – 213

Gasparini G (1998) Prognostic variables in node-negative and node-positive breast cancer. *Breast Cancer Res Treat* **52:** 321 – 331

Haapasalo JA, Nordfors KM, Hilvo M, Rantala IJ, Soini Y, Parkkila AK, Pastorekova S, Pastorek J, Parkkila SM, Haapasalo HK (2006) Expression of carbonic anhydrase IX in astrocytic tumors predicts poor prognosis. *Clin Cancer Res* **12:** 473 – 477

Hayden JA, Cote P, Bombardier C (2006) Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* **144:** 427 – 437

Hoppin JA, Tolbert PE, Taylor JA, Schroeder JC, Holly EA (2002) Potential for selection bias with tumor tissue retrieval in molecular epidemiology studies. *Ann Epidemiol* **12:** 1 – 6

Huang KC, Park DC, Ng SK, Lee JY, Ni X, Ng WC, Bandera CA, Welch WR, Berkowitz RS, Mok SC, Ng SW (2006) Selenium binding protein 1 in ovarian cancer. *Int J Cancer* **118:** 2433 – 2440

Hudis CA, Barlow WE, Costantino JP, Gray RJ, Pritchard KI, Chapman JA, Sparano JA, Hunsberger S, Enos RA, Gelber RD, Zujewski JA (2007) Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol* **25:** 2127 – 2132

Infante-Rivard C, Villeneuve JP, Esnaola S (1989) A framework for evaluating and conducting prognostic studies: an application to cirrhosis of the liver. *J Clin Epidemiol* **42:** 791 – 805

Kuijpers T, van der Windt DA, van der Heijden GJ, Bouter LM (2004) Systematic review of prognostic cohort studies on shoulder disorders. *Pain* **109:** 420 – 431

Kyzas PA, Cunha IW, Ioannidis JP (2005a) Prognostic significance of vascular endothelial growth factor immunohistochemical expression in head and neck squamous cell carcinoma: a meta-analysis. *Clin Cancer Res* **11:** 1434 – 1440

Kyzas PA, Loizou KT, Ioannidis JP (2005b) Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst* **97:** 1043 – 1055

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

Malats N, Bustos A, Nascimento CM, Fernandez F, Rivas M, Puente D, Kogevinas M, Real FX (2005) P53 as a prognostic marker for bladder cancer: a meta-analysis and review. *Lancet Oncol* **6:** 678 – 686

Marx BE, Marx M (1997) Prognosis of idiopathic membranous nephropathy: a methodologic meta-analysis. *Kidney Int* **51:** 873 – 879

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

Moher D, Schulz KF, Altman DG (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* **357:** 1191 – 1194

Pfisterer J, Kommoss F, Sauerbrei W, Renz H, du BA, Kiechle-Schwarz M, Pfleiderer A (1994) Cellular DNA content and survival in advanced ovarian carcinoma. *Cancer* **74:** 2509 – 2515

Riley RD, Abrams KR, Lambert PC, Sutton AJ, Altman DG (2006) 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) Chapter 3, pp 40 – 58. Birkhauser and Springer: 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, Sauerbrei W, Altman DG (2009) Prognostic markers in cancer: the evolution of evidence from single studies to meta-analysis, and beyond. *Br J Cancer* **100:** 1219 – 1229

Sauerbrei W (2005) Prognostic factors: confusion caused by bad quality 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. *Communs Stat Theory Methods* **35:** 1333 – 1342

Schemper M, Smith TL (1996) A note on quantifying follow-up in studies of failure time. *Control Clin Trials* **17:** 343 – 346

Scholten-Peeters GG, Verhagen AP, Bekkering GE, van der Windt DA, Barnsley L, Oostendorp RA, Hendriks EJ (2003) Prognostic factors of whiplash-associated disorders: a systematic review of prospective cohort studies. *Pain* **104:** 303 – 322

Sutcliffe P, Hummel S, Simpson E, Young T, Rees A, Wilkinson A, Hamdy F, Clarke N, Staffurth J (2009) Use of classical and novel biomarkers as prognostic risk factors for localised prostate cancer: a systematic review. *Health Technol Assess* **13:** iii, xi – iiixiii

Wadehra M, Natarajan S, Seligson DB, Williams CJ, Hummer AJ, Hedvat C, Braun J, Soslow RA (2006) Expression of epithelial membrane protein-2 is associated with endometrial adenocarcinoma of unfavorable outcome. *Cancer* **107:** 90 – 98

Wang L, Wei Q, Wang LE, Aldape KD, Cao Y, Okcu MF, Hess KR, El Zein R, Gilbert MR, Woo SY, Prabhu SS, Fuller GN, Bondy ML (2006) Survival prediction in patients with glioblastoma multiforme by human telomerase genetic variation. *J Clin Oncol* **24:** 1627 – 1632

Whiteley W, Chong WL, Sengupta A, Sandercock P (2009) Blood markers for the prognosis of ischemic stroke: a systematic review. *Stroke* **40:** e380 – e389