

BMJ Open Assessment of the extent of unpublished studies in prognostic factor research: a systematic review of p53 immunohistochemistry in bladder cancer as an example

Peggy Sekula,¹ Julia B Pressler,^{2,3} Willi Sauerbrei,¹ Peter J Goebell,⁴ Bernd J Schmitz-Dräger^{2,4}

To cite: Sekula P, Pressler JB, Sauerbrei W, *et al.* Assessment of the extent of unpublished studies in prognostic factor research: a systematic review of p53 immunohistochemistry in bladder cancer as an example. *BMJ Open* 2016;**6**: e009972. doi:10.1136/bmjopen-2015-009972

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-009972>).

PS and JBS contributed equally to this work.

Received 11 September 2015
Revised 20 June 2016
Accepted 5 July 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to

Dr Peggy Sekula;
ps@imbi.uni-freiburg.de

ABSTRACT

Objectives: When study groups fail to publish their results, a subsequent systematic review may come to incorrect conclusions when combining information only from published studies. p53 expression measured by immunohistochemistry is a potential prognostic factor in bladder cancer. Although numerous studies have been conducted, its role is still under debate. The assumption that unpublished studies too harbour evidence on this research topic leads to the question about the attributable effect when adding this information and comparing it with published data. Thus, the aim was to identify published and unpublished studies and to explore their differences potentially affecting the conclusion on its function as a prognostic biomarker.

Design: Systematic review of published and unpublished studies assessing p53 in bladder cancer in Germany between 1993 and 2007.

Results: The systematic search revealed 16 studies of which 11 (69%) have been published and 5 (31%) have not. Key reason for not publishing the results was a loss of interest of the investigators. There were no obviously larger differences between published and unpublished studies. However, a meaningful meta-analysis was not possible mainly due to the poor (ie, incomplete) reporting of study results.

Conclusions: Within this well-defined population of studies, we could provide empirical evidence for the failure of study groups to publish their results that was mainly caused by loss of interest. This fact may be coreponsible for the role of p53 as a prognostic factor still being unclear. We consider p53 and the restriction to studies in Germany as a specific example, but the critical issues are probably similar for other prognostic factors and other countries.

INTRODUCTION

In medical research, it is well accepted that a single study is unlikely to provide a definite (or final) answer to clinical questions. Common

Strengths and limitations of this study

- According to our knowledge, this is the first project which provides empirical evidence of non-publication for a specific research question due to loss of interest.
- The paradigm of p53 expression measured by immunohistochemistry might be a valid model choice to assess the extent of unpublished studies in biomarker research but may not be representative for other areas of medical research.
- Although the project is based on extensive and thorough search for published and unpublished studies in Germany, it cannot be fully excluded that relevant studies, whether published or unpublished, have been missed.
- The geographic restriction to Germany was necessary to achieve a comprehensive overview on published and unpublished studies, but this limits its general overall representativeness as does its restriction regarding the period in which studies were searched.

reasons are, for example, an insufficient study size or a limited representativeness with respect to general practice.^{1–3} Particularly, observational studies often used in prognostic research generally provide rather weak evidence because such studies are highly susceptible to various types of bias.^{3–5} Thus, it is required to accumulate evidence by systematically reviewing relevant studies and, if sensible, combining their results in a meta-analysis. Such a review including subsequent analysis may provide an overall estimate and an overview for the research community, physicians and policymakers. However, meta-analyses too are prone to bias from several sources.^{2–6} In particular, a difficult situation emerges when observational

studies compose the main source of evidence in certain research fields including prognostic marker research. Combining results from observational studies in a meta-analysis is often complicated, especially because data handling and statistical analysis frequently differ between studies. In this context, the conduct of comprehensive individual patient data (IPD) meta-analysis instead of a meta-analysis of aggregated data gained importance.^{3 7 8}

The basis of any successful review is always a systematic literature search in respective electronic databases and study registries with the aim to identify ideally all relevant studies regarding a specific question of interest. Obviously, the identification of all relevant studies worldwide is technically impossible. In addition, only randomised controlled trials have to be registered but not observational studies. Moreover, relevant studies might be missed even with an extensive search because they were not published for different reasons or at least not published until the time of the search and are thus not detectable in electronic databases or other sources.^{9 10} It was reported that over 50% of the studies in biomedical research were not fully published and thus represent avoidable waste of research evidence.¹¹ In case the search revealed only a part of the conducted studies, this may have an impact on the final conclusions—unless the identified studies that are included into a meta-analysis compose an adequate sample.

Although empirical evidence is weak, it is likely that several (observational) studies are unpublished, particularly smaller and medium-sized studies. This might especially be the case when study groups fail to publish because their results are not statistically significant. The specific problem that studies without significant results are less likely to be published has been termed as publication bias.^{12 13} The implication is that false-positive results are over-represented in the literature, most likely leading to an overestimation of the true prognostic impact.^{14–17} Although the existence of unpublished studies is well documented, the actual extent of unpublished studies in a specific setting and its effect on the derived conclusion is obviously difficult to assess because it appears impossible to appraise results of unpublished and thus invisible studies.^{13 18 19}

Since the early 1990s, p53 expression measured by immunohistochemistry (short: p53) came into focus as a potential biomarker for prognosis in patients with cancers including bladder cancer.^{20–23} Although numerous studies investigating the usefulness of p53 in bladder cancer had been conducted, its prognostic impact is still under debate, even after the conduction of several systematic reviews.^{24–28} In addition, there is clear evidence that more studies have been performed than were eventually published.^{26 29}

Two of the authors (BJSD, PJG) are actively involved in the field of marker research in bladder cancer and also have extensive long-term connections to many urological departments, pathological institutes and the respective scientific societies in Germany. Hence, it

became obvious that the topic of p53 and prognosis in bladder cancer might represent a suitable setting to investigate and assess the extent of unpublished data on this topic and to evaluate their potential for publication bias. To achieve a complete and comprehensive overview on both, published and unpublished studies, we restricted our attention to Germany and a limited time period (1993–2007). A significant number of studies originate from this area and time period.²⁴ Although such restriction to a well-defined population of studies decreases the absolute number of studies, it may still allow to estimate a combined effect unbiasedly though with increased uncertainty. A systematic literature search and the contacts of the two authors ensured the identification of nearly all studies which have been initiated.

In summary, the aim of this review was to identify all published and unpublished studies investigating the role of p53 in bladder cancer conducted over a 15-year period in Germany, to explore the differences between them and to evaluate the impact of the unpublished studies on the interpretation of p53 as prognostic biomarker in patients with bladder cancer. In addition, the reasons for not publishing results were evaluated.

METHODS

This analysis assessed studies (1) that investigated prognostic impact of p53 in bladder cancer, and (2) that were conducted in Germany through the years 1993–2007. Since no patient data were retrieved, no approval from institutional review board was obtained. The study was initiated by three of the authors in 2005 and later restarted in 2012 to finish the project.

Data retrieval

The aim of this step was to comprehensively retrieve a complete survey comprising all studies with characteristics as outlined above. To achieve this goal, an extensive, systematic search was initiated using multiple approaches and sources:

- A systematic search in Medline was conducted in 2005 and again in 2007. The search is described in figure 1. All hits were screened by BJSD and JBP and potentially relevant articles were studied.
- Assuming that this type of study is mostly performed at university hospitals, a separate literature search was launched focusing on chairmen or senior staff members from all university hospitals in Germany (BJSD, JBP).
- Assuming that urologists must be involved in this type of study, a questionnaire was sent to all urology departments in Germany through the mail server of the German Society of Urology (DGU) in 2006. Chairmen and/or senior staff members from non-responding institutions were personally contacted subsequently (BJSD).
- Programmes and conference proceedings for the study period (grey literature search) were obtained from the

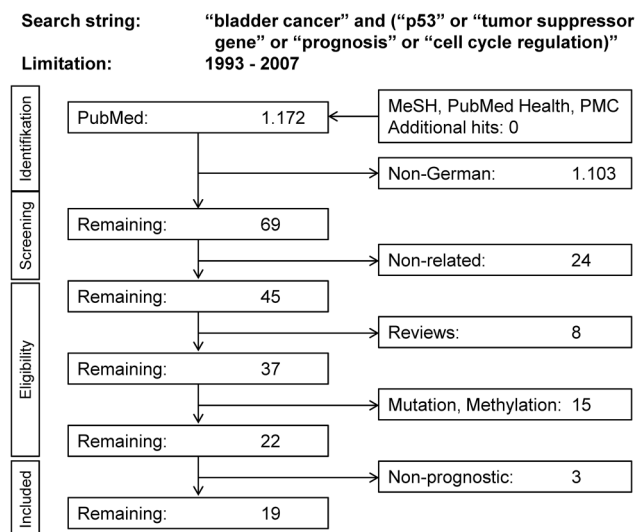


Figure 1 Flow diagram of Medline search.

German national meeting and, as far as possible, from regional scientific urological societies (BJSD, JBP).

- E. Furthermore, university hospital libraries were searched via the internet for medical theses (doctoral dissertation) studying the role of p53 in bladder cancer in 2007 (JBP).
- F. Finally, the personal knowledge of two authors (PJG, BJSD) extensively working in the field allowed to crosscheck with identified studies.

The different approaches are assumed to be sufficient to comprehensively identify all studies conducted in Germany during the chosen period.

All reports that could not be retrieved by the thorough literature search in Medline (A) were considered as 'unpublished'. If an unpublished study was identified, the researcher or the responsible faculty member was personally contacted via email to ensure that the work was still not published. If they did not respond to the email, contact was established via phone. In addition, the reason for non-publication was inquired. Obtained answers (free text) were categorised. The status of unpublished studies was checked again in May 2014.

Data extraction and analysis

From each study (published and unpublished), data for a list of items were extracted by JBP and revised by PS. Besides items on the manuscript level (reference, language, reported institution), we extracted items on the study level (type of study, recruitment period, assessed biomarkers, p53 measurement, study size, tumour stage and grade of included patients) as well as on the analysis level (assessed outcomes, number of events, statistical methods, analysis results). These items specifically reflect the research question on p53 as potential prognostic factor in bladder cancer. In the absence of a standard tool to assess study quality of observational studies, a formal assessment of study quality was not

done. However, extracted information from studies can be used as an indicator for their quality in general.

All data were analysed in a descriptive way. The original intention was to present a combined estimate for the prognostic effect of p53 on accepted end points, including tumour recurrence, disease progression and (overall or cancer-specific) survival based on published literature, and to compare it to the unpublished results to quantify potential publication bias. However, it became clear in the course of the project that it was not possible to provide any meaningful estimate, neither based on published nor based on published and unpublished data.

RESULTS

Data retrieval

The aim of the search was to identify all studies—whether published or not—on p53 investigating the prognosis of patients with bladder cancer in Germany from 1993 to 2007. The results of the Medline searches in 2005 and 2007 and the results of subsequent checks are presented in figure 1. Altogether, 19 manuscripts were selected and retrieved for data extraction.^{30–48}

To identify possibly missing additional studies, a simple questionnaire was sent to 345 urological departments in Germany in 2007 using the mail server of the German Society of Urology (DGU). In total, 13 departments were excluded because these institutions were no full urological departments but paediatric clinics, clinics for rehabilitation, or only dedicated to diagnosis. Feedback after two rounds of mailing comprised information from 192 departments (57.8%). Personal contact yielded response from another 76 institutions (22.9%). For seven departments (2.1%) no information could be obtained due to changes in staff and chairpersons. No feedback was retrieved from another 57 departments (17.2%). Altogether, information was retrieved from 268 institutions (80.7%). The feedback from university hospitals was slightly higher than from non-academic institutions (86.1% vs 80.1%, respectively).

In addition, programmes and congress proceedings for the study period were obtained from the archives of the DGU, comprising programmes of the national urological meetings, conferences on basic urological science (Experimentelle Urologie) and from 11 additional regional scientific urological societies. In summary, 65 programmes from 90 identified meetings within the relevant period (72%) could be analysed. Access to medical theses from 12 of 35 (34.3%) university libraries within Germany was obtained. Their contents were examined for potentially relevant studies.

After crosschecking with studies identified in the Medline search, five additional unpublished studies were detected through the feedback to our survey (n=2), review of medical conferences (n=3) and search for medical theses in university libraries (n=2). Two studies were identified through more than one source. There

was no other study according to the personal knowledge and communication of two of the authors (PJG, BJSJ).

Published studies

Issue: multiple reporting

Among 19 publications selected for data extraction, several manuscripts were published by the same institution. Therefore, publications were checked for independence. Publications originating from the same institution with overlapping author lists, using p53 measurements based on the same antibody and assessing patients showing a similar spectrum regarding tumour stage and tumour grade were considered as dependent.

Based on this definition, our comparison revealed 13 reports from 5 institutions with some potential overlap in study populations. Owing to insufficient information (eg, regarding period of patient recruitment), we could not assess the proportion of overlap in detail. For each of these reports, the study comprising the largest data set was selected for further analysis, whereas the remaining eight studies were excluded. In the majority of cases, this led to the most recent publication being chosen. The only exceptions hereto were the two publications from Hamburg^{37 42} and the two publications from Cologne.^{30 40} In both cases, other markers had been additionally considered in the more recent publication but only on a smaller set of patients, presumably representing a subset of the population included in the original study.

Table 1(I) summarises all 19 publications including the 8 manuscripts which were excluded because of presumed overlap in study populations.

Description of studies (n=11)

Included studies were published between 1993 and 2005. To our knowledge, the next study following this period was published in 2009 and thus not considered here.

From many reports, it remains unclear whether the study was designed and conducted in a truly prospective way or whether authors took advantage of archived specimens and patient information collected at regular follow-ups in a more or less stringent fashion. Besides diagnosis and partly specific therapy, three studies additionally requested complete follow-up data (without further specification) or follow-up data of at least 2 years for patient selection.^{39 41 47} The studies usually did not describe patient selection in detail and thus did not attempt to quantify the potential of selection bias that might arise from missing specimens or other requested data.

Information on tumour stage and tumour grade was mostly reported showing some variation between the studies (table 1(I)). While some studies focused on non-muscle invasive bladder cancer, others investigated advanced tumour stages or both. Similarly for tumour grade, some studies included patients of any tumour grade (G1–G4), while other studies focused on specific

grades (eg, only G2). Patients included in the published studies came from the corresponding single centre reporting the data. When reported, specimens were derived through transurethral resection or cystectomy.

To detect p53 overexpression by immunohistochemistry, different antibodies with different dilutions had been applied (table 2(I)). For analysis, the staining results were categorised into the two categories of negative and positive results except for one study that used a categorisation into four groups. Cut-offs varied from study to study (range for binary cut-offs: 5–40%, table 2(I)). In 7 out of 11 studies, the decisive reasons were not provided. In the remaining publications, a reference was provided as justification (n=3) or the observed median was used (n=1).

Samples sizes with respect to the prognostic analysis were often small ranging from 30 to 119 patients (median=69 patients, table 1(I)). If the number of events of interest (recurrence, progression, death) representing the effective sample size of the study was at all reported, it ranged between 11 and 63. Studies not explicitly providing the number of events reported survival rates, median survival or some other related information. The events were observed during follow-up of varying length. Reported information on the individual length of follow-up in patients ranged from 1 month (=0.1 year) to 140 months (=11.7 years). None of the published studies reported on any power or sample size calculation to justify their (effective) size of population.

Statistical analysis and results

Binary end points (recurrence or progression) were analysed in two studies using the χ^2 test.^{30 43} All other nine studies considered a time-to-event end point to assess the prognostic impact of p53: overall survival (n=2), cause-specific survival (n=2), recurrence-free survival (n=2), progression-free survival (n=4) and combined end point (tumour-free survival, n=1, cause-specific survival with preserved bladder, n=1). Four studies reported results regarding two end points,^{39 41 43 47} and two studies only presented results for subgroups.^{30 45} The definition of the time to the event was provided as duration from initial diagnosis, surgical intervention or initiation of chemotherapy to the event of interest in seven studies (78%). For statistical analysis, authors applied log-rank test (n=8) or Cox regression (univariate or multivariate; n=6) for estimation of association. A Kaplan–Meier graph was usually presented when the log-rank test was applied.

Online supplementary tables S1 and S2 contain information on reported analyses and their results of published studies. Only two of six studies that used Cox regression actually reported effect estimates in terms of HRs.^{37 47} P values were reported by 6 of 11 studies, while the other studies only provided information on significance as ‘smaller or greater than some value’ or even described the result in words (eg, ‘not significant’, ‘weakly correlated’).

Table 1 Published and unpublished studies focusing on assessment of p53 as a potential prognostic factor in bladder cancer

Reference (I) or contact person and year of analysis (II)	Reporting period*	Study size (number of events)	Tumour stage Tumour grade	Surgical treatment
(I) Published studies				
Place of reporting institution: Düsseldorf (Germany) Jankevicius <i>et al</i> (2002) ⁴⁵	NR	60 (NR)† A: 39 (NR) B: 21 (NR)	T1–T4 NR	TUR in 39 patients, cystectomy in 21 patients
Schmitz-Dräger <i>et al</i> (1997) ³⁸	NR	61 (28)	Ta–T4 G1–G3	NR
Schmitz-Dräger <i>et al</i> (1994) ³³ ‡	1985–1993	43 (7)	Ta–T4 G1–G3	NR
Place of reporting institution: Erlangen-Nürnberg and Halle (Germany) Rödel <i>et al</i> (2000) ⁴¹	1982–1996	70 (NR,NR)¶	T1–T4 G1–G4	TUR
Place of reporting institution: Hamburg (Germany) Friedrich <i>et al</i> (1997) ³⁷	NR	53 (20)	Ta–T1 G1–G3	TUR
Friedrich <i>et al</i> (2001) ⁴² ‡	NR	40 (14)	Ta, T1 G1–G3	TUR
Place of reporting institution: Hannover (Germany) Serth <i>et al</i> (1995) ³⁶	NR	69 (13)	T1 NR	TUR
Bokemeyer <i>et al</i> (1994) ³¹ ‡	NR	41 (8)	T1 NR	TUR
Kuczyk <i>et al</i> (1994) ³² ‡	NR	41 (8)	T1 G1–G2	TUR
Kuczyk <i>et al</i> (1995) ³⁵ ‡	NR	41 (8)	T1 G1–G2	TUR
Kuczyk <i>et al</i> (1995) ³⁴	NR	44 (23)	T2–T4 G2–G3	cystectomy
Place of reporting institution: Cologne (Germany) Hake <i>et al</i> (1993) ³⁰ §	NR	119 (61)† A: 40 (21) B: 79 (40)	Ta, T1–T2 G1–G2	NR
Vorreuther <i>et al</i> (1997) ⁴⁰ ‡	1978–1988	104 (53)** A: 59 (29) B: 45 (24) C: 40 (22) D: 64 (31)	Ta, T1 G1–G2	TUR
Place of reporting institution: Lübeck (Germany) Krüger <i>et al</i> (2005) ⁴⁷	1988–1998	73 (52,27)¶	T1 G2–G3	TUR
Krüger <i>et al</i> (2003) ⁴⁶ ‡	1987–1997	54 (43,19)¶	T1 G2–G3	TUR
Mahnken <i>et al</i> (2005) ⁴⁸ ‡	1987–1999	69 (48,27)¶	T1 G2–G3	TUR
Place of reporting institution: Mainz (Germany) Leissner <i>et al</i> (2001) ⁴³	1990–1997	70 (34,14)¶	Ta G2	TUR
Place of reporting institution: Mainz and Magdeburg (Germany) Wolf <i>et al</i> (2001) ⁴⁴	1989–1998	30 (11)	T1–T4 G3	TUR in 5 patients, cystectomy in 25 patients
Place of reporting institution: Münster (Germany) van Ahlen <i>et al</i> (1997) ³⁹ §	1988–1992	100 (NR,63)¶	Ta–T4 G1–G4	cystectomy

Continued

Table 1 Continued

Reference (I) or contact person and year of analysis (II)	Reporting period*	Study size (number of events)	Tumour stage Tumour grade	Surgical treatment
(II) Unpublished studies				
Place of reporting institution: Aachen (Germany) US#1: Lehnert TG (1998)	NR	121 (NR)	Ta–T4 G1–G3	NR
Place of reporting institution: Erlangen (Germany) US#2: Perez R (2001)	NR	107 (NR)	Ta, T1–T2 G1–G3	NR
Place of reporting institution: Heidelberg (Germany) US#3: Kunkel A (1996)	NR	33 (NR)	T1 G1–G3	NR
Place of reporting institution: Mainz (Germany) US#4: Adler J (1998)	NR	49 (NR)	Ta–T4 G1–G4	NR
Place of reporting institution: Saarbrücken/Homburg (Germany) US#5: Gerber M (2003)	NR	115 (NR)	T1–T4 G2–G4	NR

*Reporting period usually refers to the time period of patient inclusion.

†Analysis was only done within two non-overlapping subgroups.

‡Associated (excluded) studies with the respective included study mentioned in row above.

¶Studies assessing two end points.

§Publication in German language.

**Analysis was only done within partly overlapping subgroups (A and B as well as C and D add up to whole study population).

NR not reported; TUR, transurethral resection.

Owing to the heterogeneity in study populations, different outcomes assessed and different statistical methods applied, but, in particular, because of insufficient information on effect estimates or p values, a meaningful meta-analysis cannot be conducted. Although it might be possible to derive crude estimates and p values based on reported numbers for some of the studies, this kind of approach is (often) not sensible for observational studies when effect estimates cannot be adjusted for the presence of confounding factors.

Table 3(I) presents an overview regarding the final conclusion on the significance of the impact of the p53 measures on the given end point. Considering all 17 reported analyses of the 11 included studies simultaneously, more than half of them (59%) revealed non-significant findings. Results show a similarly ambiguous pattern, even when restricting analyses that included 69 (observed median of overall population size) or more patients (effective samples size of 13–63 events; data not shown). Studies reporting ‘significant’ results, however, consistently provide a trend towards a worse outcome for patients with higher accumulated p53 measured by immunohistochemistry.

Finally, when comparing results among the five institutions publishing more than once, the reports per institution were generally consistent regarding their conclusion.

Overall quality of studies

As many items that we consider relevant in this setting are incompletely/not reported by the authors of these

studies, the reporting quality is generally poor. As a consequence, the quality of these studies cannot be objectively assessed and thus remains questionable.

Unpublished studies in comparison

A total of 5 unpublished studies were detected, corresponding to 31% of the 16 initiated studies (table 1(II)). Explanations provided for not publishing the results were obtained for four of the five studies (80%) and were loss of interest of the investigator and/or change in staff. Figure 2 presents the distribution in time of unpublished studies (year of presentation or thesis) compared with the published studies (year of publication) assuming that unpublished studies would have been published in the same year or at least not significantly later.

Comparing the sizes of unpublished to published studies, the pool of unpublished studies contains three studies (US#1:Lehnert TG 1998; US#2:Perez R 2001; US#5:Gerber M 2003) that are of similar size as the largest published study (table 1).³⁰ However, it is unclear whether all patients of unpublished studies did actually enter the prognostic analysis or whether this represents the overall pool of, for example, available biopsy results. Other study characteristics of unpublished studies such as tumour stage and tumour grade were not remarkably different from published studies by showing some variation (table 1). The same holds true for antibodies and cut-offs used as far as reported (table 2).

Two studies assessed tumour recurrence, two assessed progression and one assessed both. Altogether, results

Table 2 p53 antibodies for immunohistochemistry

Antibody	Reference (I) or contact person and year of analysis (II)	Dilution	Cut-off to define positivity
(I) Published studies			
AB-6	Jankevicius <i>et al</i> (2002) ⁴⁵	1:100	20%
BP5312	Hake <i>et al</i> (1993) ³⁰	NR	10%
CM1	van Ahlen <i>et al</i> (1997) ³⁹	NR	<20%, 30–80%, >80%*
DO-1	Schmitz-Draeger <i>et al</i> (1997) ³⁸	1:100	5%
	Wolf <i>et al</i> (2001) ⁴⁴	1:100	20%
DO-7	Rödel <i>et al</i> (2000) ⁴¹	1:50	10%
	Friedrich <i>et al</i> (1997) ³⁷	1:100	5%
	Leissner <i>et al</i> (2001) ⁴³	1:100	20%
pAB1801	Serth <i>et al</i> (1995) ³⁶	1:50	20%
	Kuczyk <i>et al</i> (1995) ³⁴	1:50	40%
	Krüger <i>et al</i> (2005) ⁴⁷	1:50	7% (observed median)
(II) Unpublished studies			
DO-1	US#1:Lehnert TG (1998)	1:100	<10%, 10–20%, >20%
DO-1 and pAB1801	US#4:Adler J (1998)	NR	DO-1: 9%, pAB1801: 7.2% (observed medians)
	US#5:Gerber M (2003)	NR	various cut-offs used; eg, 40%
unknown	US#2:Perez R (2001)	NR	various cut-offs used; eg, 25%
	US#3:Kunkel A (1996)	NR	<20%, 20–80%, >80%

*Presumably printing mistake in source publication.
NR: not reported.

Table 3 Final conclusion based on reported analyses of time-to-event or binary outcomes

		Final conclusion: significant?	
Outcome	Studies (n)	Yes	No
(I) Published studies	11		
Overall survival	3	0	3
Cancer-specific survival	2	1	1
Recurrence-free survival	2	1	1
Progression-free survival	4	2	2
Combined end point	2	1	1
Binary end point	4	2	2
(recurrence, progression)			
Total number of available results	17	7	10
(II) Unpublished studies	5		
Recurrence	4	2	2
Progression	2	0	2
Combined end point	1	0	1
Total number of available results	7	2	5

If several results for same outcome using same set of patients within a study were available, conclusion was drawn from the analysis that is presumably least biased (eg, multivariable Cox regression).

from seven analyses were reported. Similar to published studies, only p values or the study's conclusions are available. Two of the seven reported analyses presented significant results (29%), while all other comparisons yielded insignificant results (table 3). See online supplementary table S3 for a detailed overview. Remarkably,

one of the significant results was obtained in a study that also assessed the prognostic impact of p53 in further separate analyses not revealing any further significant association (US#4:Adler J 1998). There is no information available on how the association between outcome and p53 was statistically assessed.

Overall, collected information is similarly incomplete as for published studies. The study quality is thus difficult to assess and questionable.

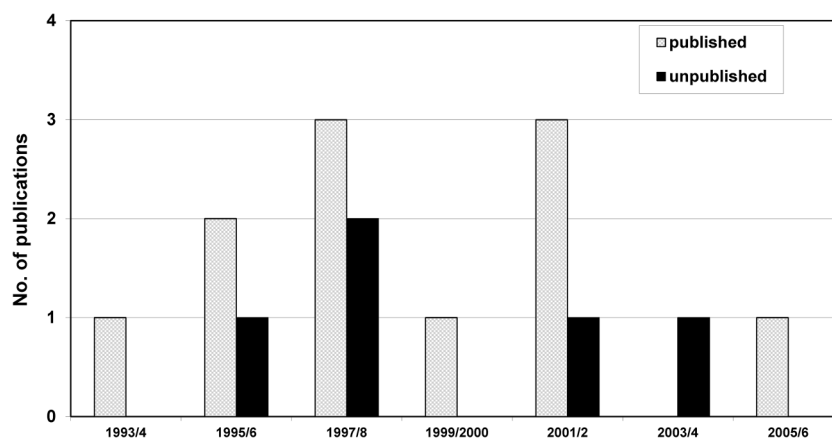
DISCUSSION

Non-publication due to loss of interest

Since the presence of publication bias may lead to incorrect interpretation and conclusions of available data/evidence, it is necessary to evaluate the extent and possible impact of unpublished studies. In this project, we assessed this problem in the paradigm of p53 as a potential prognostic factor in bladder cancer in a well-defined area (all centres in Germany conducting research on the question of interest) over a representative period of time (1993–2007). Our approach enabled us to illustrate this issue in a very specific setting—maybe for the first time.

Our extensive search for published and unpublished studies revealed 16 studies assessing the prognostic impact of p53 in bladder cancer of which 11 (69%) have been published and 5 (31%) have not. Thus, conclusions based on published data alone are prone to publication bias because the omission of the publication most likely occurred for some reason and not randomly. In our project, reasons were loss of interest of the investigators and/or change in staff. At least to some extent, both might be related to the conclusion that the study results were thought to be insufficiently innovative or

Figure 2 Published and unpublished studies over time. For published studies, the year of publication is presented; for unpublished studies, the year of presentation at a congress (or similar) or the year of dissertation is presented. We assume that publication of unpublished results would have been not much later but at least within our considered time frame (up to 2007).



not convincing. As ‘change in staff’ may also be related to the loss of interest of the research group or may imply the change of scientific focus within the research group, we decided to term it non-publication due to ‘loss of interest’ in general. Similar to previous findings, it seems to mainly reflect the investigator’s (in)decision rather than other circumstances such as rejection of the manuscript by a journal.^{9 49}

The paradigm of p53 as a potential prognostic factor in bladder cancer appears to be a suitable example to illustrate this issue. This setting is especially eligible because of the historical development of marker research during this period and the restricted number of possible sites, leading to the feasibility to conduct a retrospective study. Furthermore, these studies are characterised by a similar design and similar end points. In addition, surgery and follow-up is conducted by urologists ensuring a close contact to participants of such studies. Starting rather promising in the early 1990s, p53 was the subject of many studies in the last 25 years. Still, the issue of its prognostic impact in bladder cancer is not settled even after the conduction of meta-analyses.^{24 26–28} In the course of the years, some evidence that more studies have been performed than were eventually published was accumulated.^{26 29} It may therefore be speculated that this loss of interest might have been intensified over the years due to the contradictory results published.

While this project has several intriguing aspects, there are also some limitations to be discussed: first, although this project is based on extensive and thorough search for published and unpublished studies including pre-existing knowledge of these investigations by two authors (PJG, BJSD), we cannot fully exclude that we may have missed studies whether published or not. Regarding published studies, our search strategy in only one electronic data source (Medline) might not have revealed all relevant studies. However, the impact of further data sources on search result was recently shown to be modest.⁵⁰ Moreover, the literature search was restricted to the period from 1993 to 2007 and was not updated later on because of the complexity to search for unpublished studies. As a consequence, we cannot present

information on the current situation in this field. Misclassification of published studies as unpublished was ruled out by personally contacting responsible study group members inquiring the publication status. In addition, we may have missed some unpublished studies due to the incomplete feedback from contacted people and institutions, incomplete acquisition of abstract books as well as limited access to university hospital libraries.

Another limitation may be the geographic restriction to Germany questioning the project’s representativeness on a larger scale. This step, however, was necessary because a European or even worldwide search of this character is impossible. In order to meet the primary interest of a complete identification of published and unpublished studies in one region rather than on general representativeness, this approach appears justified. Although the restriction of a review to such a well-defined population of studies decreases the number of studies overall and thus increases the uncertainty when estimating a combined effect, it also has several methodological advantages. Identifying all studies in a well-defined population might allow to derive a (nearly) unbiased estimate and it might offer the option to conduct IPD meta-analysis, a key issue in a meta-analysis of observational studies. This idea was already proposed by Altman in 1983.⁵¹

Finally, the paradigm of p53 might be a valid model choice to assess the extent of unpublished studies and their impact in biomarker research but may not be representative for other areas of medical research.

Non-publication and publication bias in medical research

Non-publication of study results is a general issue in medical research. Chalmers and Glasziou stated in 2009 that over 50% of the studies in biomedical research were not fully published and thus represent avoidable waste of research evidence.¹¹ As consequence, published results might be affected by publication bias leading to an overestimation of the true effect.^{12–15} In prognostic cancer research, a respective survey revealed that nearly all included articles reported significant results.¹⁷

Unfortunately, the described history of p53 is not uncommon in the field of biomarker research. Despite

the postulated clinical usefulness of biomarkers in general, most biomarkers do not enter clinical practice giving rise to several publications discussing this situation and evaluating the reasons.^{52–55} Since marker research and especially immunohistochemistry appear to be simple to perform and interpret and guidelines on this type of research are still not implemented, it is also assumed that non-publication occurs more frequently than, for example, in the area of randomised clinical trials assessing therapeutic effects.⁵⁶

Past studies assessing non-publication applied several approaches to identify unpublished studies: one indirect way to assess it was to compare results of scientific reports to non-scientific communications such as newspaper articles.¹⁴ A more direct way is the approach to identify study groups or single people working in the field using large databases such as Medline or via related publications or with the help of membership lists of related organisations and, subsequently, to directly ask investigators whether they have knowledge of unpublished studies.^{29 57} Although this approach is rather straightforward, some unpublished studies conducted, for example, by single researchers or smaller research groups (new or less active in the field and thus not present in the databases) might be missed. Another approach is the identification and the evaluation of the status of studies indicated to ethics committees, review boards, funding bodies or trial registries.^{58–61} Although registries will become more exhaustive in the future, those sources are momentarily very limited in their content of observational studies. Moreover, the completeness of such sources may not be reliably available for any area of medical research.

In order to circumvent the majority of the caveats mentioned before, we used a composed approach to identify unpublished studies addressing the same specific research question in Germany. On the one hand, we used several of the methods applied in the past by taking advantage of different sources including the search in Medline as well as programmes and congress proceedings to identify people and study groups working in the field. The list was then further extended by directly addressing urologic departments and an investigation of university libraries. Moreover, we used the knowledge of two of us (PJJ, BJSD) about the research field and the general research structure in Germany. Altogether, for p53 as a potential prognostic marker in bladder cancer, our strategy provides a comprehensive overview of published and unpublished studies in the given time period in Germany. The described method may provide a new transparent template to perform this type of research.

Can we summarise the effect of p53?

Since the discussion on the relevance of p53 as a prognostic marker in bladder cancer is still under debate, the question arises whether unpublished studies would change the current understanding and interpretation of

the data.^{26 27} One of the initial aims was to identify all relevant additional studies (including their evidence) and to evaluate their attributable effect on the evidence we have from the interpretation of published results. While we were successful with regard to the retrieving of the completeness of p53 work in this period in Germany, we failed to gather sufficient valid and comparable (individual patient) data for the anticipated comparison. As a meaningful meta-analysis of the obtained data is not possible, the current study cannot address the impact of unpublished studies on the interpretation of the evidence of p53 as a prognostic marker. However, since published and unpublished studies do not show larger differences in their characteristics and results, it could be speculated that there would be no general change in the judgement of p53 as prognostic marker if unpublished results had been reported and thus implemented in a comprehensive analysis.

Reasons for not being able to provide a meaningful combined estimate in the current study are manifold: first, there is the issue that studies are poorly reported which makes the assessment of study quality and interpretation of results difficult and sometimes impossible. This includes a poor description of the study and an insufficient report of the conducted analysis and its results. For example, only two of the six published studies that used Cox regression in their analysis reported effect estimates. Thus, a meta-analysis based on effect estimates is not meaningful. Alternatively, one may try to combine reported p values. But even this kind of analysis—of questionable value—is not feasible because most of the authors did not list p values for assessed end points. Poor reporting of study results, however, is a general issue in medical research and not limited to the presented work. For cancer prognostic studies empirical evidence has been provided.⁶² Second, the large heterogeneity between studies questions the usefulness of a combined estimate in general. This conclusion is based on observed differences in patient populations, methodology (eg, usage of different antibodies in different dilutions, different cut-offs), end points and statistical methods.

Another aspect often overlooked is that studies are frequently too small to provide any reliable result.²⁶ The general issue of small sample sizes of studies was very well illustrated in a review on tumour markers in neuroblastoma.⁶³ The identified studies of our project are no exception. The population sizes of published and unpublished studies vary between 30 and 119 patients. The number of events, which reflects the effective sample size in studies with a survival outcome, was even much smaller with values ranging from 11 to 63. None of the included studies provided rationale for sample sizes. Consequently, the power of studies is assumed to be limited. Subgroup analyses are not sensible. Unfortunately, the largest of the published studies included (overall 119 patients, 61 events) only reports subgroup analyses.³⁰

Improvement of biomarker research

In general, our project on p53 illustrates quite impressively the necessity to improve the general situation in prognostic research regarding design and conduction of studies including their statistical analysis and reporting. As there is a long way until a biomarker may reach clinical use, different aspects have to be considered depending on the phase of biomarker research; standards must be developed and applied.^{64–66}

Fortunately, the last decade has seen several important contributions, in particular regarding reporting health research. Following the example of the Consolidated Standards of Reporting Trials (CONSORT) guidelines for randomised controlled trials, reporting guidelines have been developed for many types of observational studies.⁶⁷ For tumour marker studies, the Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) guideline including a 20-items checklist was introduced in 2005 and detailed explanations and elaborations of these items were published later.^{68–69} Good reporting is an issue which could be easily improved by following the guidelines, but many more issues are relevant and some of them are either discussed controversially or are a challenge for research. The PROGNosis REsearch Strategy (PROGRESS) group (<http://progress-partnership.org/>) published a series of articles to provide a framework on different aspects in prognostic research.^{70–73}

In addition, the large proportion of unpublished studies detected in our project also underlines the necessity of a preregistration of any study, whether interventional or observational.^{3 7 74–77} Prospective registration of all studies helps not only to quantify the extent of unpublished (or not yet published) studies. It was even demonstrated that it helps to improve study design and methodology.⁷⁸ Moreover, published protocols may allow an assessment of studies included into a systematic review regarding risk of bias in a more sufficient way.⁷⁹ Currently, study registration is only obligatory for interventional trials and optional for observational studies, the usual design in prognostic research.

Also concerning suitable statistical analyses, substantial contributions for its improvement have been made recently. Unfortunately, many of these developments are ignored in practice and methods with known weaknesses are still used.^{56–69} Aiming to derive guidance documents for key issues of the analysis of observational studies, the STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative (<http://www.stratos-initiative.org/>) was founded recently.⁸⁰

Finally, to derive a summary assessment for a marker of interest, it is obvious that a systematic review and a meta-analysis are required.⁷ Whereas IPD meta-analyses of prognostic factors studies have been the exception 10 years ago, several collaborative groups have impressively shown that IPD meta-analyses are possible and that such projects may provide deeper insight into the role of a prognostic factor of interest.^{3 81} However, when

starting such a review project, many aspects, especially regarding the definition of well-defined population of studies, need to be considered.

Conclusion

Altogether, our strategy provides a comprehensive overview of published and unpublished studies on p53 as a potential prognostic marker in bladder cancer in the given time period in Germany. The described method may provide a new transparent template to perform this type of research.

Using a well-defined cohort of studies, we could provide empirical evidence of non-publication mainly caused by loss of interest of the investigators. This survey suggests that about 30% of the studies were not published despite the fact that there were no larger differences between published and unpublished studies.

We consider p53 as a specific example, but the critical issues of non-publication and bad reporting are very similar for research investigating prognostic factors in other diseases. Briefly, we point to some recent initiatives which have been started to improve on this frustrating situation.

Author affiliations

¹Institute for Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg, Freiburg, Germany

²Department of Urology, Schön-Klinik Nürnberg Fürth, Fürth, Germany

³KUNO University Children's Hospital, Regensburg, Germany

⁴Department of Urology, University Clinic of Erlangen, Waldkrankenhaus St. Marien, Erlangen, Germany

Acknowledgements The authors thank Martin Gerber (Homburg), Peter Effert (Aachen) and Markus Hohenfellner (Heidelberg) for their support in retrieving unpublished material. JBP contributed to the project in fulfilment of the requirements for obtaining the degree 'Dr med.' at the Friedrich-Alexander University, Erlangen-Nürnberg, Germany. The article processing charge was funded by the German Research Foundation (DFG) and the Albert Ludwigs University of Freiburg in the funding programme Open Access Publishing.

Contributors JBP, PJG and BJSD contributed to the conception and design of the study as well as to the acquisition of data. PS and WFS were responsible for the analysis of acquired data and formatting of the results. All authors were involved in the interpretation of the results and in the drafting of the manuscript. The version for publication was approved by all of them.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement This project is based on extracted data from published and unpublished studies that can be requested from the corresponding author.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Sutton AJ, Abrams KR, Jones DR, *et al.* *Methods for meta-analysis in medical research*. Chichester: John Wiley & Sons Ltd, Wiley Series in Probability and Statistics, 2000.

2. Egger M, Smith GD, Altman DG. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Publishing Group, 2001.
3. Riley RD, Sauerbrei W, Altman DG. Prognostic markers in cancer: the evolution of evidence from single studies to meta-analysis, and beyond. *Br J Cancer* 2009;100:1219–29.
4. Guyatt GH, Oxman AD, Kunz R, et al. What is “quality of evidence” and why is it important to clinicians? *BMJ* 2008;336:995–8.
5. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;126:376–80.
6. Roberts I, Ker K, Edwards P, et al. The knowledge system underpinning healthcare is not fit for purpose and must change. *BMJ* 2015;350:h2463.
7. Sauerbrei W, Holländer N, Riley RD, et al. Evidence-based assessment and application of prognostic markers: the long way from single studies to meta-analysis. *Commun Stat-Theor M* 2006;35:1333–42.
8. Thompson S, Kaptoge S, White I, et al. Statistical methods for the time-to-event analysis of individual participant data from multiple epidemiological studies. *Int J Epidemiol* 2010;39:1345–59.
9. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992;267:374–8.
10. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997;315:640–5.
11. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;374:86–9.
12. Montori VM, Smieja M, Guyatt GH. Publication bias: a brief review for clinicians. *Mayo Clin Proc* 2000;75:1284–8.
13. Song F, Eastwood AJ, Gilbody S, et al. Publication and related biases. *Health Technol Assess* 2000;4:1–115.
14. Hernandez JF, Mantel-Teeuwisse AK, van Thiel GJ, et al. Publication trends in newspapers and scientific journals for SSRIs and suicidality: a systematic longitudinal study. *BMJ Open* 2011;1:e000290.
15. Kicinski M. How does under-reporting of negative and inconclusive results affect the false-positive rate in meta-analysis? A simulation study. *BMJ Open* 2014;4:e004831.
16. Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:696–701.
17. Kyzas PA, Denaxa-Kyza D, Ioannidis JPA. Almost all articles on cancer prognostic markers report statistically significant results. *Eur J Cancer* 2007;43:2559–79.
18. Dickersin K. The Existence of Publication Bias and Risk-Factors for Its Occurrence. *JAMA* 1990;263:1385–9.
19. Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE* 2008;3:e3081.
20. Sarkis AS, Dalbagni G, Cordoncardo C, et al. Nuclear overexpression of p53-protein in transitional cell bladder-carcinoma—a marker for disease progression. *J Natl Cancer Inst* 1993;85:53–9.
21. Cordon-Cardo C, Dalbagni G, Saez GT, et al. p53 mutations in human bladder cancer: genotypic versus phenotypic patterns. *Int J Cancer* 1994;56:347–53.
22. Dowell SP, Hall PA. The p53 tumor-suppressor gene and tumor prognosis—is there a relationship? *J Pathol* 1995;177:221–4.
23. Uchida T, Wada C, Ishida H, et al. p53 mutations and prognosis in bladder-tumors. *J Urology* 1995;153:1097–104.
24. Schmitz-Dräger BJ, Goebell PJ, Ebert T, et al. p53 immunohistochemistry as a prognostic marker in bladder cancer. Playground for urology scientists? *Eur Urol* 2000;38:691–9; discussion 700.
25. Schmitz-Dräger BJ, Goebell PJ, Heydthausen M. p53 immunohistochemistry in bladder cancer. Combined analysis: a way to go? *Urol Oncol* 2000;5:204–10.
26. Malats W, Bustos A, Nascimento CM, et al. p53 as a prognostic marker for bladder cancer: a meta-analysis and review. *Lancet Oncol* 2005;6:678–86.
27. Goebell PJ, Groshen SG, Schmitz-Dräger BJ. International Study-Initiative on Bladder C. p53 immunohistochemistry in bladder cancer—a new approach to an old question. *Urol Oncol* 2010;28:377–88.
28. Patel T, Pitman M, McKiernan JM. Bladder cancer: a review of clinical management and prognostic factors. *Minerva Urol Nefrol* 2010;62:377–85.
29. Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst* 2005;97:1043–55.
30. Hake R, Vorreuther R, Borchmann P, et al. [Monoclonal antibodies (MIB 1, PC 10, 486p and p53) as prognostic factors for recurrent urothelial carcinoma of the urinary bladder]. *Verh Dtsch Ges Pathol* 1993;77:236–40.
31. Bokemeyer C, Kuczyk MA, Serth J. Clinical implications of the p53 tumor-suppressor gene. *N Engl J Med* 1994;330:865–5.
32. Kuczyk MA, Serth J, Hervatin C, et al. Detection of p53 tumor-suppressor-gene protein in bladder tumors and prostate cancer: possible clinical implications. *World J Urol* 1994;12:345–51.
33. Schmitz-Dräger BJ, Vanroeyen CRC, Grimm MO, et al. p53 accumulation in precursor lesions and early stages of bladder-cancer. *World J Urol* 1994;12:79–83.
34. Kuczyk MA, Bokemeyer C, Serth J, et al. p53 overexpression as a prognostic factor for advanced stage bladder cancer. *Eur J Cancer* 1995;31A:2243–7.
35. Kuczyk MA, Serth J, Bokemeyer C, et al. [Value of the proliferation status (PCNA) and p53 immunohistochemistry as a prognostic factor for the clinical course of superficial cancer of the urinary bladder]. *Urologe A* 1995;34:146–52.
36. Serth J, Kuczyk MA, Bokemeyer C, et al. p53 immunohistochemistry as an independent prognostic factor for superficial transitional cell carcinoma of the bladder. *Br J Cancer* 1995;71:201–5.
37. Friedrich MG, Schwaibold H, Wintzer O, et al. p53 in noncancerous bladder mucosa as a marker of disease recurrence in patients with superficial transitional cell carcinoma of the bladder. *Urol Oncol* 1997;3:125–31.
38. Schmitz-Dräger BJ, Kushima M, Goebell P, et al. p53 and MDM2 in the development and progression of bladder cancer. *Eur Urol* 1997;32:487–93.
39. van Ahlen H, Wülfing C, Roth S, et al. Prognostische Wertigkeit des p53-Tumor-Suppressor-Gens beim urothelien Harnblasenkarzinom. *Aktuelle Urologie* 1997;28:216–22.
40. Vorreuther R, Hake R, Borchmann P, et al. Expression of immunohistochemical markers (PCNA, Ki-67, 486p and p53) on paraffin sections and their relation to the recurrence rate of superficial bladder tumors. *Urol Int* 1997;59:88–94.
41. Rödel C, Grabenbauer GG, Rödel F, et al. Apoptosis, p53, bcl-2, and Ki-67 in invasive bladder carcinoma: possible predictors for response to radiochemotherapy and successful bladder preservation. *Int J Radiat Oncol* 2000;46:1213–21.
42. Friedrich MG, Riethdorf S, Erbersdobler A, et al. Relevance of p53 gene alterations for tumor recurrence in patients with superficial transitional cell carcinoma of the bladder. *Eur Urol* 2001;39:159–66.
43. Leissner J, Hohenfellner R, Thuroff JW, et al. Prognostic significance of histopathological grading and immunoreactivity for p53 and p21/WAF1 in grade 2 pTa transitional cell carcinoma of the urinary bladder. *Eur Urol* 2001;39:438–45.
44. Wolf HK, Stober C, Hohenfellner R, et al. Prognostic value of p53, p21/WAF1, Bcl-2, Bax, Bak and Ki-67 immunoreactivity in pT1 G3 urothelial bladder carcinomas. *Tumor Biol* 2001;22:328–36.
45. Jankevicius F, Goebell P, Kushima M, et al. p21 and p53 immunostaining and survival following systemic chemotherapy for urothelial cancer. *Urol Int* 2002;69:174–80.
46. Krüger S, Thorns C, Stocker W, et al. Prognostic value of MCM2 immunoreactivity in stage T1 transitional cell carcinoma of the bladder. *Eur Urol* 2003;43:138–45.
47. Krüger S, Lange I, Kausch I, et al. Protein expression and gene copy number analysis of topoisomerase 2alpha, HER2 and p53 in minimally invasive urothelial carcinoma of the urinary bladder—a multitissue array study with prognostic implications. *Anticancer Res* 2005;25:263–71.
48. Mahnken A, Kausch I, Feller AC, et al. E-cadherin immunoreactivity correlates with recurrence and progression of minimally invasive transitional cell carcinomas of the urinary bladder. *Oncol Rep* 2005;14:1065–70.
49. Easterbrook PJ, Berlin JA, Gopalan R, et al. Publication bias in clinical research. *Lancet* 1991;337:867–72.
50. Halladay CW, Trikalinos TA, Schmid IT, et al. Using data sources beyond PubMed has a modest impact on the results of systematic reviews of therapeutic interventions. *J Clin Epidemiol* 2015;68:1076–84.
51. Altman DG. Evaluating a series of clinical trials of the same treatment. The James Lind Library. <http://www.jameslindlibrary.org/altman-dg-1981/>
52. McShane LM, Altman DG, Sauerbrei W. Identification of clinically useful cancer prognostic factors: what are we missing? *J Natl Cancer I* 2005;97:1023–5.
53. Diamandis EP. The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? *BMC Med* 2012;10.
54. Kern SE. Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. *Cancer Res* 2012;72:6097–101.

55. Ioannidis JP. Biomarker failures. *Clin Chem* 2013;59:202–4.
56. Sauerbrei W. Prognostic factors. Confusion caused by bad quality design, analysis and reporting of many studies. *Adv Otorhinolaryngol* 2005;62:184–200.
57. Shadish WR, Doherty M, Montgomery LM. How many studies are in the file drawer? an estimate from the family marital psychotherapy literature. *Clin Psychol Rev* 1989;9:589–603.
58. Decullier E, Chapuis F. Impact of funding on biomedical research: a retrospective cohort study. *BMC Public Health* 2006;6:165.
59. Jones CW, Handler L, Crowell KE, et al. Non-publication of large randomized clinical trials: cross sectional analysis. *BMJ* 2013;347:f6104.
60. Turner S, Wright D, Maeso R, et al. Publication rate for funded studies from a major UK health research funder: a cohort study. *BMJ Open* 2013;3:pii: e002521.
61. Blumle A, Meerpohl JJ, Schumacher M, et al. Fate of clinical research studies after ethical approval--follow-up of study protocols until publication. *PLoS ONE* 2014;9:e87184.
62. Mallett S, Royston P, Dutton S, et al. Reporting methods in studies developing prognostic models in cancer: a review. *BMC Med* 2010;8:20.
63. Riley RD, Heney D, Jones DR, et al. A systematic review of molecular and biological tumor markers in neuroblastoma. *Clin Cancer Res* 2004;10:4–12.
64. Hayes DF, Bast RC, Desch CE, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst* 1996;88:1456–66.
65. Pepe MS, Etzioni R, Feng ZD, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* 2001;93:1054–61.
66. Goebell PJ, Groshen SL, Schmitz-Drager BJ. Guidelines for development of diagnostic markers in bladder cancer. *World J Urol* 2008;26:5–11.
67. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637–9.
68. McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;93:387–91.
69. Altman DG, McShane LM, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC Med* 2012;10:51.
70. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595.
71. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med* 2013;10:e1001380.
72. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10.
73. Hingorani AD, Windt DA, Riley RD, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ* 2013;346:e5793.
74. Dickersin K. How important is publication bias? A synthesis of available data. *Aids Educ Prev* 1997;9:15–21.
75. Peat G, Riley RD, Croft P, et al. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. *PLoS Med* 2014;11:e1001671.
76. Andre F, McShane LM, Michiels S, et al. Biomarker studies: a call for a comprehensive biomarker study registry. *Nat Rev Clin Oncol* 2011;8:171–6.
77. Altman DG. The time has come to register diagnostic and prognostic research. *Clin Chem* 2014;60:580–2.
78. Liu JP, Han M, Li XX, et al. Prospective registration, bias risk and outcome-reporting bias in randomised clinical trials of traditional Chinese medicine: an empirical methodological study. *BMJ Open* 2013;3:pii: e002968.
79. Jefferson T, Jones MA, Doshi P, et al. Risk of bias in industry-funded oseltamivir trials: comparison of core reports versus full clinical study reports. *BMJ Open* 2014;4:e005253.
80. Sauerbrei W, Abrahamowicz M, Altman DG, et al. STRENGTHENING Analytical Thinking for Observational Studies: the STRATOS initiative. *Stat Med* 2014;33:5413–32.
81. Ioannidis JP, Rosenberg PS, Goedert JJ, et al. International Meta-analysis of HIVHG. Commentary: meta-analysis of individual participants' data in genetic epidemiology. *Am J Epidemiol* 2002;156:204–10.