

# Did the reporting of prognostic studies of tumour markers improve since the introduction of REMARK guideline?

Sekula, Peggy; Mallett, Susan; Altman, Douglas; Sauerbrie, Willi

*License:*  
Creative Commons: Attribution (CC BY)

*Document Version*  
Peer reviewed version

*Citation for published version (Harvard):*  
Sekula, P, Mallett, S, Altman, D & Sauerbrie, W 2017, 'Did the reporting of prognostic studies of tumour markers improve since the introduction of REMARK guideline? A comparison of reporting in published articles', *PLoS ONE*, vol. 12, no. 6, PONE-D-17-00107R1.

[Link to publication on Research at Birmingham portal](#)

## General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

## Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

1 **Did the reporting of prognostic studies of tumour markers**

2 **improve since the introduction of REMARK guideline?**

3 **A comparison of reporting in published articles.**

4 **Short title:** Evaluation of reporting quality of prognostic studies of tumour markers

5

6 Peggy Sekula<sup>1\*</sup>, Susan Mallett<sup>2</sup>, Douglas G Altman<sup>3</sup>, Willi Sauerbrei<sup>1</sup>

7

8 1 Institute for Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg,

9 Freiburg, Germany

10 2 Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham, UK

11 3 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal

12 Sciences, University of Oxford, Oxford, United Kingdom

13

14 \* Corresponding author

15 E-Mail: ps@imbi.uni-freiburg.de (PS)

# 1 **Abstract**

2 Although biomarkers are perceived as highly relevant for future clinical practice, few biomarkers  
3 reach clinical utility for several reasons. Among them, poor reporting of studies is one of the major  
4 problems. To aid improvement, reporting guidelines like REMARK for tumour marker prognostic  
5 (TMP) studies were introduced several years ago. The aims of this project were to assess whether  
6 reporting quality of TMP-studies improved in comparison to a previously conducted study assessing  
7 reporting quality of TMP-studies (PRE-study) and to assess whether articles citing REMARK (citing  
8 group) are better reported, in comparison to articles not citing REMARK (not-citing group).

9 For the POST-study, recent articles citing and not citing REMARK (53 each) were identified in selected  
10 journals through systematic literature search and evaluated in same way as in the PRE-study. Ten of  
11 the 20 items of the REMARK checklist were evaluated and used to define an overall score of  
12 reporting quality.

13 The observed overall scores were 53.4% (range: 10%-90%) for the PRE-study, 57.7% (range: 20%-  
14 100%) for the not-citing group and 58.1% (range: 30%-100%) for the citing group of the POST-study.

15 While there is no difference between the two groups of the POST-study, the POST-study shows a  
16 slight but not relevant improvement in reporting relative to the PRE-study. Not all the articles of the  
17 citing group, cited REMARK appropriately. Irrespective of whether REMARK was cited, the overall  
18 score was slightly higher for articles published in journals requesting adherence to REMARK than for  
19 those published in journals not requesting it: 59.9% versus 51.9%, respectively.

20 Several years after the introduction of REMARK, many key items of TMP-studies are still very poorly  
21 reported. A combined effort is needed from authors, editors, reviewers and methodologists to  
22 improve the current situation. Good reporting is not just nice to have but is essential for any research  
23 to be useful.

# 1 Introduction

2 Major advances in molecular biology and in analytical laboratory methods including new (high-  
3 throughput) technologies have enabled the detection and the measurement of a wide range of  
4 biomarkers in the human body. This has led to an increasing number of studies assessing the utility of  
5 biomarkers in a medical context [1, 2]. In this regard, a biomarker is an objectively measured  
6 characteristic with biological, clinical, genetic, histological or pathological background [3]. Biomarker  
7 measurement can be based on a single assessment or on a combination of information from several  
8 assessments (e.g. scores) [1, 4]. Biomarkers are already used successfully and routinely in different  
9 areas of medicine (e.g. serum creatinine to assess kidney function [5, 6]) and are perceived as highly  
10 relevant for future clinical practice using stratified or personalized medicine, where biomarkers may  
11 be useful to assist medical decision making, ideally underpinned by recommendations in clinical  
12 guidelines. Areas of biomarker use include but are not restricted to different aspects of patient care  
13 [2, 4, 7, 8]:

- 14 • screening of people to allow early detection of diseases,
- 15 • differential diagnosis of patients,
- 16 • stratification of patients for treatments,
- 17 • monitoring of treatment response and treatment compliance, and
- 18 • identification of risk groups related to patients' prognosis.

19 Biomarkers are also useful in the discovery and development of new treatments, through their role  
20 in elucidation of disease processes [4, 8]. Additionally, biomarkers are important in the design of  
21 studies and trials, allowing stratification of participants and use as surrogate endpoints [8, 9].

22 There are several important steps to establish the clinical value of a particular biomarker,  
23 including well-designed and well-reported clinical studies [1, 9-11]. Yet very few biomarkers have  
24 established clinical value [1, 12, 13], as exemplified by cancer research where it is estimated that  
25 fewer than 1% of biomarkers originally proposed as important have entered clinical practice [14].  
26 Researchers have investigated reasons for this unsatisfactory situation. Different types of failures

1 were distinguished, especially where results of subsequent studies contradict preceding study results  
2 [14-16]. There are major concerns that the poor quality of studies can lead to misleading results and  
3 consequently mistaken claims of utility [1, 3, 12, 17, 18].

4         As biomarker studies can be challenging, methodologists have highlighted the need for more  
5 transparency, standardization and harmonization to improve studies [11, 13, 19-23]. Overall, this will  
6 not only help to improve quality of individual studies but also enhance the ability to compare results  
7 between studies – a prerequisite for evidence synthesis and meta-analysis. A typical example is  
8 provided by p53. Since the early 1990s, p53 has been measured by immunohistochemistry and  
9 assessed as a potential prognostic biomarker in bladder cancer in many studies. Although  
10 researchers invested a lot of effort, time and money, the research question is still unanswered [24-  
11 28]. This situation is a consequence of many different methodological issues, such as small study  
12 populations and variation in study methods resulting in differences in the handling of measurements  
13 (e.g. different cutpoints used to define positive biomarker results).

14         Poor reporting is another major problem in these studies. Many biomarker studies are never  
15 reported at all and there is evidence that publication is linked to study results; Kyzas *et al* found that  
16 <1.5% of published prognostic marker studies were found to have only “negative” results [29].  
17 Within published studies, there are problems with the selective reporting of results and with the  
18 poor quality of reporting of methods and results [3, 21]. For tumour marker prognostic studies (TMP-  
19 studies), evidence for bad reporting has been provided [30]. In general, publications that are of poor  
20 quality can be essentially considered as a waste of research resources [31]. Worse, poor reporting in  
21 published studies might lead to incorrect conclusions about the evidence relating to a specific  
22 question.

23         To help overcome issues regarding the poor quality of reporting, guidelines for specific  
24 research areas were introduced. A valuable research hub is provided by the EQUATOR Network  
25 providing searchable access to reporting guidelines appropriate for many study designs and specific  
26 study features [32]. Among others, the REMARK guideline (short: REMARK) is a reporting guideline

1 specifically for TMP-studies assessing biomarkers in relation to future health outcomes in cancer  
2 patients. This guideline was published in seven journals in 2005/6 [33-39]. For convenience, the  
3 authors provided a checklist of 20 items addressing different parts of a manuscript. REMARK can be  
4 used by authors, editors and reviewers to check the reporting quality of a study report (**S1 Doc**). In  
5 addition, an extensive 'Explanation and Elaboration' (E&E) article was published in 2012, providing  
6 detailed information and examples of good reporting practice for each of these checklist items [40,  
7 41]. The need for REMARK was supported by a study that strikingly showed the poor reporting  
8 quality of 50 TMP-studies published in 2006-7 [30]. Because of the usual delay before an article is  
9 published, it is most unlikely that the authors of the assessed articles knew REMARK at the time of  
10 writing their manuscript (pre-REMARK period).

11         The aim of this project was to evaluate whether the quality of reporting of such studies has  
12 improved since the publication of REMARK (post-REMARK period). We repeated the previous study  
13 (short: PRE-study) using articles published between 2007 to 2012 (short: POST-study) using methods  
14 and definitions as similar as possible, to allow a fair comparison with previous findings [30]. Some  
15 TMP-studies cite the REMARK guidelines demonstrating awareness of REMARK, sometimes because  
16 journals like *Breast Cancer Research and Treatment (BCRT)* ask for adherence to REMARK in  
17 submitted manuscripts [42]. In contrast, authors of articles not citing the guideline are more likely to  
18 be unaware of REMARK or may not be using the checklist. In this study, we also addressed the  
19 question whether citing the REMARK guideline or not is related to the reporting quality. In summary,  
20 the two aims of the project are:

- 21             1. Has there been any improvement in reporting quality since introduction of REMARK?
- 22             2. Is reporting better in studies citing REMARK?

23

24

# 1 Material and Methods

2 Because only published data from studies in humans were utilized, no approval from an ethic  
3 committee was obtained.

4 To allow a direct comparison to the previous work [30], the POST-study was designed in a  
5 very similar way (choice of journals, study selection, data extraction). In this study, two groups of  
6 publications were distinguished: (A) publications that cited one of the seven REMARK publications  
7 (citing group) and (B) publications that did not cite REMARK (not-citing group) [33-39]. Similarly to  
8 the PRE-study, it was planned to include 50 articles per group as sufficient size in this methodological  
9 study to address questions of interest.

10

## 11 1 Literature search

12 To identify TMP-studies citing REMARK, a literature search was done in Web of Science in March  
13 2013. References of all publications citing at least one of the REMARK publications were extracted  
14 and imported into Endnote [33-39]. After removal of duplicates (n=72), 998 articles published in 278  
15 different journals were identified. Among them, 134 publications were in one of the five previously  
16 considered cancer journals: *Cancer [Canc]*, *Cancer Research [CaRes]*, *International Journal of Cancer*  
17 *[IJC]*, *Journal of Clinical Oncology [JCO]*, *Clinical Cancer Research [CCR]*.

18 The 134 identified articles published in the five journals considered in the PRE-study were  
19 then examined to identify for each journal the 10 most recent TMP-studies that cited REMARK. A  
20 detailed description of the eligibility criteria can be found in the **S2 Doc**. Essentially, studies assessing  
21 the prognostic impact of a specific biomarker on an outcome of clinical importance (e.g. cancer-  
22 specific survival) in cancer patients were eligible. The search revealed 10 articles each from *JCO* and  
23 *CCR*, 7 from *IJC*, 6 from *Canc* and 1 from *CaRes*. Because of this result, we decided to exclude *CaRes*  
24 from further consideration and to include two further cancer journals (*Breast Cancer Research and*  
25 *Treatment [BCRT]* and *British Journal of Cancer [BJC]*) for which 10 articles each could be identified.

1 Altogether, the citing group comprised 53 articles. Although about 80% of the included manuscripts  
2 were published in 2011 and 2012, a few dated back to 2007.

3 To identify publications not citing REMARK, we aimed to obtain for each article citing  
4 REMARK another article from the same journal that did not, closely matched in time (publication  
5 year and, if possible, issue). The same number of articles (n=53) was identified forming the not-citing  
6 group.

7 The described search is depicted in **Fig 1**. The references of all selected articles are listed in  
8 **S3 Doc**.

9

10 **Fig 1: Literature search – flow chart**

11

## 12 **2 Data extraction**

13 For all 106 articles from the six journals we obtained the full text. For data extraction, we used the  
14 same standardized form that had been used in the PRE-study (**S4 Doc**) [30]. This form lists several  
15 elements (specific questions) addressing different items of the REMARK checklist. The focus of data  
16 extraction led on information related to methods and results of a study. Because of the general  
17 character of each checklist item, a specific item is often described by more than 1 element of the  
18 data extraction form.

19 To ensure good comparability of extracted data with past results, a pilot data extraction for  
20 eight articles was done in duplicate by the author (SM) who mainly did the data extraction in the  
21 PRE-study and another author (PS) who was responsible for it in the POST-study. Results of these  
22 extractions were compared and differences clarified before data extractions were done for the  
23 remaining articles by PS alone.

24 For articles in which several biomarkers were assessed in a study in parallel, the biomarker  
25 first mentioned in title or abstract for which a multivariable analysis was done was defined to be the  
26 focus of the data extraction. A similar approach was used when different study populations were



1 assessed within a single article. Two groups of time-to-event outcomes were distinguished: death-  
2 related outcomes (overall survival, cancer-specific survival) and other time-to-event outcomes  
3 (disease-free survival, time until recurrence/relapse). Similarly, when several outcomes were  
4 assessed in a study the data extraction focused on the outcome that was first mentioned in title or  
5 abstract for which a multivariable analysis had been conducted.

6 Importantly, this project focuses only on the assessment of reporting quality and not on the  
7 general appropriateness of methods, including study design, assessed biomarkers, statistical  
8 methods and outcomes considered.

9

### 10 **3 Analyses**

11 We addressed our first aim on the improvement over time by comparing the results obtained in the  
12 PRE-study to those of the not-citing group. The second aim on difference in reporting when citing or  
13 not citing REMARK was addressed by comparing the results for the citing and not-citing groups within  
14 the current POST-study.

15 The intended comparisons were descriptively conducted with respect to 10 of the 20 items of  
16 the REMARK checklist that are related to methods and results of a manuscript (**Table 1**). For each  
17 article, we evaluated whether information for each item was provided (yes/no) by combining  
18 extracted information of elements assigned to that item. Details regarding selection of checklist  
19 items and definitions how items were evaluated are provided in **S5 Doc**. Finally, for each article an  
20 overall score was obtained as the percentage of items addressed out of 10.

21

22

1 **Table 1: Overview of the 10 assessed items of the REMARK checklist**

No.	Manuscript part	Item of REMARK checklist	Short description	Abbreviation used in article
1	Methods	2	<b>PAT</b> ient characteristics	PAT
2		6	Study <b>DES</b> ign: patient selection & time period	DES
3		7	Clinical <b>END</b> points	END
4		9	Rationale for sample <b>SIZ</b> e	SIZ
5		10	All statistical <b>METH</b> ods	MET
6	Results	12	<b>FLO</b> w of patients	FLO
7		13	Distribution of <b>DEM</b> ographic characteristics	DEM
8		14	<b>REL</b> ationship between marker and standard variables	REL
9		15	<b>UNI</b> variate analyses	UNI
10		16	<b>MULT</b> ivariable analyses	MUL

2

3           Only 10 checklist items were included in the assessment of adherence to REMARK as we used  
 4 only items we could assess objectively and that could be assessed on TMP-studies from any research  
 5 area. Items 1, 19 and 20 referring to the introduction and the discussion of an article were  
 6 considered too subjective and require subject-specific expert knowledge, and so had not been  
 7 included in the data extraction form that was already used in the PRE-study. Similarly, the seven  
 8 items 3, 4, 5, 8, 11, 17 and 18 referring to the methods and the results of an article were excluded  
 9 because their evaluation essentially requires profound expert knowledge with respect to the medical  
 10 background and methodology. For more details, see **S5 Doc**.

11

12

## 1 **4 Reporting**

2 This study assesses the reporting quality of published TMP-studies. For such a ‘research on research’-  
3 project, no specific reporting guideline is available. The current project, however, shows some  
4 features (observational kind, literature search) that allow us to use different reporting statements as  
5 guidance. Specifically, we used the STROBE guideline for general aspects of the project and the  
6 PRISMA statement for aspects around literature search [43, 44].

7

8

# 1 Results

2

## 3 1 Selected journals and assessed articles

### 4 Overall

5 The POST-study was planned to use the same journals as much as was feasible, but some changes to  
 6 journals was required for practical reasons (**Table 2**). All journals are of higher impact (impact factor  
 7 2012 >4). Three journals (*BCRT*, *BJC*, *JCO*) belong to the group of journals that have published  
 8 REMARK. These three journals and *CCR* explicitly ask authors submitting a manuscript for adherence  
 9 to REMARK (**Table 2**).

10

11 **Table 2: Cancer journals included in the PRE-study and in the POST-study**

Journal (alphabetical order)	PRE-study		POST-study		Impact factor <sup>†</sup> 2012	Publication of REMARK	Author instructed to adhere to REMARK	
		N assessed articles		N assessed articles <sup>*</sup>			02/2009 <sup>‡</sup>	08/2014
<i>BCRT</i>	-	-	✓	10/10	4.5	YES [38]	YES [42]	YES
<i>BJC</i>	-	-	✓	10/10	5.1	YES [33]	UNK	YES
<i>Canc</i>	✓	10	✓	6/6	5.2	NO	NO	NO
<i>CaRes</i>	✓	10	-	- <sup>§</sup>	8.6	NO	NO	NO
<i>CCR</i>	✓	10	✓	10/10	7.8	NO	YES	YES
<i>IJC</i>	✓	10	✓	7/7	6.2	NO	NO	NO
<i>JCO</i>	✓	10	✓	10/10	18.0	YES [35]	YES	YES

12 <sup>\*</sup> N citing group/n not-citing group; <sup>†</sup> source: InCites™ Journal Citation Reports; <sup>‡</sup> check was done within the PRE-study; <sup>§</sup> journal was

13 excluded because only one eligible article citing REMARK was identified; UNK=unknown

14

1           The sample included 53 articles in both the citing group and the not-citing group (total  
2 n=106). Similarly to the PRE-study, the distribution of cancer sites was diverse. As a consequence of  
3 the additional inclusion of *BCRT*, however, the proportion of breast and/or ovarian cancer studies  
4 was higher in the current sample (PRE-study: 30%, POST-study: 44%). Articles in the not-citing and  
5 citing groups were well matched by journal, year and issue (**S1 Table**).

6

## 7 **Citing group**

8           At least one of the REMARK publications was referenced in all the articles assigned to the citing  
9 group. Since REMARK is a methodological tool, its citation is expected to be given in the methods  
10 section of the article, with a statement like “The study is reported in accordance to the REMARK  
11 guideline”. Although REMARK was indeed cited most often in the methods section (n=39, 74%), some  
12 citations appeared in other sections of the manuscripts. The statements in which REMARK was cited  
13 varied greatly. While some authors correctly referred to the reporting of the study, other authors  
14 referred to REMARK in relation to the design, the conduct and the analysis of the study. For example,  
15 the statement “*This analysis was conducted according to the reporting recommendations for tumor  
16 marker guidelines for prognostic studies ...*” was provided by the authors in the methods section [45].  
17 Other statements are difficult to understand, such as “*Protein expression was evaluated using a  
18 semiquantitative weighted histoscore method by two observers as previously described ... in  
19 accordance with the Reporting Recommendations for Tumour Marker Prognostic Studies (REMARK)  
20 guidelines ...*” [46].

21           Three manuscripts were accompanied by a completed REMARK checklist [47-49]; two of  
22 these had overlapping authorship [47, 48]. For unknown reasons, none of these lists cover the full  
23 REMARK checklist of 20 items. Moreover, some explanations were difficult to assess, for example  
24 regarding the item ‘Flow of patients’, the authors stated “*This is not a staged analysis. The evaluated  
25 cohort is described ...*” [48].

26

## 1 **2 Comparison of reporting quality**

### 2 **Not-citing group (POST-study) versus PRE-study**

3 Overall, there was a slight but not relevant improvement in the mean overall score: PRE-study 53.4%  
4 (range: 10%-90%), not-citing group of POST-study 57.7% (range: 20%-100%, Wilcoxon rank sum test:  
5  $p = 0.33$ , **Fig 2**). This small difference, however, vanished when we included only articles published in  
6 the four journals assessed in both periods: PRE-study 56.5% (range: 10%-80%,  $n=40$ ), not-citing group  
7 of POST-study 56.4% (range: 20%-80%,  $n=33$ ). Some items showed an improvement in reporting from  
8 the past to the present, while others showed a decline (**S2 Table**). An improvement, for example, was  
9 visible for item 2 'Patient characteristics' (PAT): PRE-study 54%, POST-study 72%. In this case, the  
10 improvement was also visible for the single assigned elements like the element 'Selection of patients'  
11 (PAT1) showing improvement from 64% in the PRE-study to 77% in the not-citing group (**Fig 3A**). In  
12 contrast, a remarkable decline from past to present was seen for item 9 'Rationale for sample size'  
13 (SIZ; **Fig 3B, S5 Doc**). Overall, there remains much room for improvement of reporting.

14

15 **Fig 2: Percentages of articles adequately reporting information for 10 selected items of the**  
16 **REMARK checklist.**

17 The items are sorted by percentages obtained in the PRE-study [30]. See Table 1 or S2 Table for  
18 explanation of abbreviations used for different checklist items.

19

20 **Fig 3: Percentages of articles adequately reporting information for two checklist items and their**  
21 **single elements respectively assigned.**

22 (A) Checklist item 2 'Patient characteristics' (PAT), (B) checklist item 9 'Rationale for sample size'  
23 (SIZ). See S2 Table for explanation of abbreviations used for different elements of data extraction and  
24 checklist items.

25

## 1 **Citing group versus not-citing group (POST-study)**

2 When comparing the not-citing group and the citing group, there was essentially no difference in  
3 mean scores: not-citing group 57.7% (range: 20%-100%), citing group 58.1% (range: 30%-100%, **Fig**  
4 **2**). Again, some single checklist items showed an improvement in reporting from the past to the  
5 present, while others showed a decline. Most pronounced, item 7 'Clinical endpoints' (END) was  
6 reported better in the citing group than in the non-citing group (40% vs 66%, respectively; **S2 Table**),  
7 whereas it was the other way around for item 13 'Distribution of demographic characteristics' (DEM;  
8 55% vs 42%). **Fig 3** similarly illustrates observed percentages for item 2 (PAT) and item 9 (SIZ).

9

## 10 **Additional analysis**

11 Because we observed some unexpected statements by authors citing REMARK which could imply a  
12 lack of understanding of REMARK as a reporting guideline, an additional comparison was made of  
13 articles grouped by journals requesting (4 journals, 80 articles) or not requesting (2 journals, 26  
14 articles) adherence to REMARK, irrespective whether authors cited or not cited REMARK (**Table 2**).  
15 On average, the overall score for articles published in journals requesting adherence to REMARK was  
16 higher (59.9%) than for the other group (51.9%). This ordering was also present for each single  
17 checklist item.

18

19

# 1 Discussion

2 Several years after REMARK was introduced, and with many published discussions of the reporting  
3 quality in health research and prominence given to the role of poor reporting in contributing to  
4 research waste, some improvement of the reporting quality of TMP-studies was expected [30, 50,  
5 51]. However, our assessment of articles from the post-REMARK period did not reveal any relevant  
6 improvement over the quality of articles assessed in the earlier study [30]. The overall reporting  
7 quality is still very poor. Authors still frequently fail to report important aspects of their study such as  
8 the source of the study population, fully defined clinical endpoints, and an explanation of the sample  
9 size.

10 Moreover, we observed essentially no difference in reporting quality when comparing  
11 articles citing and not citing REMARK. Because citing REMARK means the author of the respective  
12 article is aware of the guideline, one would expect to see superior reporting quality compared to  
13 articles not citing REMARK. Our findings, however, raise the question of whether the main scope of  
14 REMARK is really understood. To overcome any misunderstanding the REMARK group already  
15 published a manuscript that elaborates and explains each item of the REMARK checklist in detail [40,  
16 41]. However, authors of articles assessed in this project (published  $\leq 2012$ ) could not have known  
17 this amendment because it was published in 2012.

18 Because of these disappointing results we decided to conduct an additional unplanned  
19 comparison between reporting qualities of articles published in journals requesting or not requesting  
20 adherence to REMARK in the submission guidelines. This revealed somewhat better reporting in the  
21 group of articles published in journals requesting adherence to REMARK.

22

23



## 1 **1 Limitations of study**

2 To allow a fair comparison of results between past and current assessments, the current project was  
3 designed to be as similar as possible to the previous study. Also, the current team largely overlaps  
4 with the team of the past study. Furthermore, all the documents including the data extraction form  
5 could be utilized. A pilot study was conducted to ensure comparability between data extractions in  
6 the past and current projects. Still, some systematic differences between the two surveys cannot be  
7 ruled out. In addition although judged as sufficient to address the methodological research question,  
8 the number of studies assessed was relatively small in both the pre-study and the current study.

9 One obvious limitation of this study is that we could not identify enough articles in all  
10 journals considered in the first study, so two new journals (*BCRT*, *BJC*) were included. Since both  
11 additional journals published REMARK and requesting adherence to it, the overall result might be  
12 biased. For this reason, an additional analysis was conducted by restricting articles to those published  
13 in the four journals *Canc*, *CCR*, *IJC* and *JCO* that were considered in both assessments. As result, the  
14 small improvement observed in the overall sample vanished. Overall we found no improvement in  
15 reporting quality of prognostic factor studies in the first (about six) years since REMARK was  
16 published. Repeating the investigation with papers published after more than ten years (say in 2016)  
17 may provide better results.

18 Another issue relates to the overall score used to evaluate quality of reporting. The overall  
19 score included summation of sufficiently reported REMARK items, often based on several elements  
20 of the data extraction form. For transparency, a description of the overall score and detailed results  
21 are provided in the supporting information (**S5 Doc**, **S2 Table**).

22

23

## 1 **2 Our findings in the context of published literature**

2 To our knowledge, there is just one other published study assessing quality of TMP-studies, which  
3 reviews studies of prognostic markers for colorectal cancer published in 2009-11, a slightly earlier  
4 period to the current project [52]. The authors assessed adherence to the complete REMARK  
5 checklist and found a mean score of 60 out of 78, but still emphasize deficiencies in reporting similar  
6 to those seen in our study across all cancers.

7 Concern about reporting quality applies across all areas of health research. To overcome this  
8 problem reporting guidelines for many different study designs and research areas are available [32].  
9 Similarly to our project, other study groups also assessed the question of whether reporting quality  
10 improved over time. For randomized controlled trials and in relation to the CONSORT statement,  
11 modest improvement in reporting quality was reported but reporting was still considered suboptimal  
12 [53, 54]. For other guidelines like STARD or STROBE, some slight improvements were also reported  
13 [55-58]. The current study on REMARK is essentially in line with those other reported results.

14 Da Costa *et al* systematically examined reasons for citing STROBE guideline [43, 59]. Similar  
15 to our observations, the authors reported that the guideline is often used inappropriately. These  
16 observations raise doubts on the general understanding of reporting guidelines and their aim, as  
17 already discussed in 2008 [60].

18 Evidence of a relation between reporting quality and endorsement of reporting guidelines by  
19 journals is limited [54, 61]. Our data suggest that a request of adherence by the journal might be  
20 useful. In order to provide conclusive evidence, well-planned prospective studies in cooperation with  
21 editors are needed to explore and enhance journal editor led interventions to improve reporting  
22 [61].

23 Based on our experience in the current project, we became aware that expert knowledge of  
24 the research subject and methods is often required to evaluate details needed for good reporting.  
25 Editors and reviewers may find it hard to recruit and focus experts on reporting as well as results of  
26 research studies. For authors writing a manuscript, access to sufficient expertise should be easier

1 because the research team should include experts relevant to the clinical and methodological aspects  
2 of a study. On the other hand, reporting guidelines are misunderstood by many authors [62], and  
3 further initiatives like the E&E paper for REMARK may be very helpful [40, 41].

4

### 5 **3 Quality of medical research in general**

6 In general, the quality of medical research, including other aspects besides reporting, has been  
7 criticized heavily in the last years [1, 14, 19, 31, 63-66]. To overcome these issues, several important  
8 contributions as well as the introduction of reporting guidelines have been seen recently. For  
9 example, the PROGnosis RESearch Strategy (PROGRESS) group published a series of articles to  
10 provide a framework on different aspects in prognostic research [11, 67-70]. Also, the STRATOS  
11 (STRengthening Analytical Thinking for Observational Studies) initiative was founded recently that  
12 aims to derive guidance documents related to design and analysis of observational studies [71, 72].

13 Overall, the need for transparency in medical research still appears to lack widespread  
14 acceptance and research endeavour [21, 23, 73]. Researchers remain insufficiently aware of the need  
15 to make their research clear and understandable to other researchers, as well as practising  
16 physicians, patients and other stakeholders (e.g. pharmaceutical companies, funding agencies).  
17 Particularly in medical research it is important that studies can be repeated by other research groups,  
18 requiring transparency through good reporting of research methods and results.

19 Registration of all studies and data sharing [1, 23, 73-76] have been recommended to  
20 improve knowledge of ongoing and past research. In this context, good reporting is a main  
21 prerequisite. Even a well-conducted and well-analysed study that is poorly reported can be  
22 considered as waste of resources.

23

24

## 1 **Conclusions**

2 Tumour marker prognostic studies are still very poorly reported. To improve the situation the  
3 REMARK recommendations need to be followed. However, this study is another example illustrating  
4 that publication of guidelines is insufficient and that more pressure on authors, reviewers and editors  
5 is needed to improve on this unfortunate situation. We support the proposal of one reviewer of this  
6 manuscript that an electronic checklist (a web-based form of the checklist on which the authors can  
7 indicate where in the manuscript information for an item is addressed) can be a valuable instrument  
8 of the submission process. Ideally, such an electronic document can also provide further information  
9 about the reporting items. We hope that more journals will be willing to request such checklists in  
10 their submission process. Good reporting is not just nice to have. It is essential for any research to be  
11 useful but also for the limitations of research to be understood. Good reporting is also essential for  
12 systematic reviews that bring together and overview research studies to achieve a high level of  
13 evidence.

# 1 **Acknowledgement**

2 We thank Saskia Motschall, Clemens Wachter and Corinna Roller (student helpers) for their support  
3 during the literature search, article acquisition and selection of published studies.

4 The article processing charge was funded by the German Research Foundation (DFG) and the  
5 University of Freiburg in the funding programme Open Access Publishing.

6

# 1   **References**

- 2   1. Riley RD, Sauerbrei W, Altman DG. Prognostic markers in cancer: the evolution of evidence from  
3   single studies to meta-analysis, and beyond. *Br J Cancer*. 2009;100(8):1219-29.  
4   doi:10.1038/sj.bjc.6604999.
- 5   2. Rifai N, Watson ID, Miller WG. Commercial immunoassays in biomarkers studies: researchers  
6   beware! *Clin Chem*. 2012;58(10):1387-8. doi:10.1373/clinchem.2012.192351.
- 7   3. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Jones DR, Heney D et al. Reporting of prognostic  
8   markers: current problems and development of guidelines for evidence-based practice in the future.  
9   *Br J Cancer*. 2003;88(8):1191-8. doi:10.1038/sj.bjc.6600886.
- 10   4. Anderson DC, Kodukula K. Biomarkers in pharmacology and drug discovery. *Biochem Pharmacol*.  
11   2014;87(1):172-88. doi:10.1016/j.bcp.2013.08.026.
- 12   5. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI et al. A new equation to  
13   estimate glomerular filtration rate. *Ann Intern Med*. 2009;150(9):604-12.
- 14   6. Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T et al. Estimating glomerular  
15   filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-9.  
16   doi:10.1056/NEJMoa1114248.
- 17   7. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive  
18   biomarkers. *J Natl Cancer Inst*. 2009;101(21):1446-52. doi:10.1093/jnci/djp335.
- 19   8. Micheel CM, Nass SJ, Omenn GS. Evolution of Translational Omics: Lessons Learned and the Path  
20   Forward. National Academies Press. 2012. doi:DOI: 10.17226/13297.
- 21   9. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer*.  
22   1994;69(6):979-85.
- 23   10. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M et al. Phases of biomarker  
24   development for early detection of cancer. *J Natl Cancer Inst*. 2001;93(14):1054-61.

- 1 11. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA et al. Prognosis Research  
2 Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med.* 2013;10(2):e1001380.  
3 doi:10.1371/journal.pmed.1001380.
- 4 12. McShane LM, Altman DG, Sauerbrei W. Identification of clinically useful cancer prognostic  
5 factors: what are we missing? *J Natl Cancer Inst.* 2005;97(14):1023-5. doi:10.1093/jnci/dji193.
- 6 13. Plebani M. Lessons from controversy: biomarkers evaluation. *Clin Chem Lab Med.*  
7 2013;51(2):247-8. doi:10.1515/cclm-2013-0012.
- 8 14. Kern SE. Why your new cancer biomarker may never work: recurrent patterns and remarkable  
9 diversity in biomarker failures. *Cancer Res.* 2012;72(23):6097-101. doi:10.1158/0008-5472.CAN-12-  
10 3232.
- 11 15. Ioannidis JP. Evolution and translation of research findings: from bench to where? *PLoS Clin*  
12 *Trials.* 2006;1(7):e36. doi:10.1371/journal.pctr.0010036.
- 13 16. Ioannidis JP. Biomarker failures. *Clin Chem.* 2013;59(1):202-4.  
14 doi:10.1373/clinchem.2012.185801.
- 15 17. Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in  
16 breast cancer. *Breast Cancer Res Treat.* 1998;52(1-3):289-303.
- 17 18. Sauerbrei W. Prognostic factors. Confusion caused by bad quality design, analysis and reporting  
18 of many studies. *Adv Otorhinolaryngol.* 2005;62:184-200. doi:10.1159/000082508.
- 19 19. Altman DG. The scandal of poor medical research. *BMJ.* 1994;308(6924):283-4.
- 20 20. Ioannidis JP. Genetic and molecular epidemiology. *J Epidemiol Community Health.*  
21 2007;61(9):757-8. doi:10.1136/jech.2006.059055.
- 22 21. McShane LM, Hayes DF. Publication of tumor marker research results: the necessity for complete  
23 and transparent reporting. *J Clin Oncol.* 2012;30(34):4223-32. doi:10.1200/JCO.2012.42.6858.
- 24 22. Behrens T, Bonberg N, Casjens S, Pesch B, Bruning T. A practical guide to epidemiological practice  
25 and standards in the identification and validation of diagnostic markers using a bladder cancer  
26 example. *Biochim Biophys Acta.* 2014;1844(1 Pt A):145-55. doi:10.1016/j.bbapap.2013.07.018.

- 1 23. Peat G, Riley RD, Croft P, Morley KI, Kyzas PA, Moons KG et al. Improving the transparency of  
2 prognosis research: the role of reporting, data sharing, registration, and protocols. *PLoS Med*.  
3 2014;11(7):e1001671. doi:10.1371/journal.pmed.1001671.
- 4 24. Schmitz-Drager BJ, Goebell PJ, Ebert T, Fradet Y. p53 immunohistochemistry as a prognostic  
5 marker in bladder cancer. Playground for urology scientists? *Eur Urol*. 2000;38(6):691-9;discussion  
6 700. doi:20364.
- 7 25. Malats N, Bustos A, Nascimento CM, Fernandez F, Rivas M, Puente D et al. P53 as a prognostic  
8 marker for bladder cancer: a meta-analysis and review. *Lancet Oncol*. 2005;6(9):678-86.  
9 doi:10.1016/S1470-2045(05)70315-6.
- 10 26. Goebell PJ, Groshen SG, Schmitz-Drager BJ, International Study-Initiative on Bladder C. p53  
11 immunohistochemistry in bladder cancer--a new approach to an old question. *Urol Oncol*.  
12 2010;28(4):377-88. doi:10.1016/j.urolonc.2010.03.021.
- 13 27. Kamat AM, Hegarty PK, Gee JR, Clark PE, Svatek RS, Hegarty N et al. ICUD-EAU International  
14 Consultation on Bladder Cancer 2012: Screening, diagnosis, and molecular markers. *Eur Urol*.  
15 2013;63(1):4-15. doi:10.1016/j.eururo.2012.09.057.
- 16 28. Sanguedolce F, Bufo P, Carrieri G, Cormio L. Predictive markers in bladder cancer: do we have  
17 molecular markers ready for clinical use? *Crit Rev Clin Lab Sci*. 2014;51(5):291-304.  
18 doi:10.3109/10408363.2014.930412.
- 19 29. Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Almost all articles on cancer prognostic markers report  
20 statistically significant results. *Eur J Cancer*. 2007;43(17):2559-79. doi:10.1016/j.ejca.2007.08.030.
- 21 30. Mallett S, Timmer A, Sauerbrei W, Altman DG. Reporting of prognostic studies of tumour  
22 markers: a review of published articles in relation to REMARK guidelines. *Br J Cancer*.  
23 2010;102(1):173-80. doi:10.1038/sj.bjc.6605462.
- 24 31. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP et al. Biomedical research:  
25 increasing value, reducing waste. *Lancet*. 2014;383(9912):101-4. doi:10.1016/S0140-6736(13)62329-  
26 6.



- 1 32. EQUATOR Network. <http://www.equator-network.org>.
- 2 33. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. REporting  
3 recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer. 2005;93(4):387-91.  
4 doi:10.1038/sj.bjc.6602678.
- 5 34. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. REporting  
6 recommendations for tumour MARKer prognostic studies (REMARK). Eur J Cancer. 2005;41(12):1690-  
7 6. doi:10.1016/j.ejca.2005.03.032.
- 8 35. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. Reporting  
9 recommendations for tumor marker prognostic studies. J Clin Oncol. 2005;23(36):9067-72.  
10 doi:10.1200/JCO.2004.01.0454.
- 11 36. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. Reporting  
12 recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst.  
13 2005;97(16):1180-4. doi:10.1093/jnci/dji237.
- 14 37. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. REporting  
15 recommendations for tumor MARKer prognostic studies (REMARK). Nat Clin Pract Oncol.  
16 2005;2(8):416-22.
- 17 38. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. REporting  
18 recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat.  
19 2006;100(2):229-35. doi:10.1007/s10549-006-9242-8.
- 20 39. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. Reporting  
21 recommendations for tumor marker prognostic studies (remark). Exp Oncol. 2006;28(2):99-105.
- 22 40. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting recommendations for tumor marker  
23 prognostic studies (REMARK): explanation and elaboration. BMC Med. 2012;10:51.  
24 doi:10.1186/1741-7015-10-51.

- 1 41. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker  
2 Prognostic Studies (REMARK): explanation and elaboration. *PLoS Med.* 2012;9(5):e1001216.  
3 doi:10.1371/journal.pmed.1001216.
- 4 42. Hayes DF, Ethier S, Lippman ME. New guidelines for reporting of tumor marker studies in breast  
5 cancer research and treatment: REMARK. *Breast Cancer Res Treat.* 2006;100(2):237-8.  
6 doi:10.1007/s10549-006-9253-5.
- 7 43. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP et al. The  
8 Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement:  
9 guidelines for reporting observational studies. *Ann Intern Med.* 2007;147(8):573-7.
- 10 44. Moher D, Liberati A, Tetzlaff J, Altman DG, Prisma Group. Preferred reporting items for  
11 systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097.  
12 doi:10.1371/journal.pmed.1000097.
- 13 45. Debled M, MacGrogan G, Brouste V, Mathoulin-Pelissier S, Durand M, Mauriac L. Prognostic  
14 factors of early distant recurrence in hormone receptor-positive, postmenopausal breast cancer  
15 patients receiving adjuvant tamoxifen therapy: results of a retrospective analysis. *Cancer.*  
16 2007;109(11):2197-204. doi:10.1002/cncr.22667.
- 17 46. Spears M, Pederson HC, Lyttle N, Gray C, Quintayo MA, Brogan L et al. Expression of activated  
18 type I receptor tyrosine kinases in early breast cancer. *Breast Cancer Res Treat.* 2012;134(2):701-8.  
19 doi:10.1007/s10549-012-2076-7.
- 20 47. Larsson A, Johansson ME, Wangefjord S, Gaber A, Nodin B, Kucharzewska P et al. Overexpression  
21 of podocalyxin-like protein is an independent factor of poor prognosis in colorectal cancer. *Br J*  
22 *Cancer.* 2011;105(5):666-72. doi:10.1038/bjc.2011.295.
- 23 48. Eberhard J, Gaber A, Wangefjord S, Nodin B, Uhlen M, Ericson Lindquist K et al. A cohort study of  
24 the prognostic and treatment predictive value of SATB2 expression in colorectal cancer. *Br J Cancer.*  
25 2012;106(5):931-8. doi:10.1038/bjc.2012.34.

- 1 49. Maniecki MB, Etzerodt A, Ulhoi BP, Steiniche T, Borre M, Dyrskjot L et al. Tumor-promoting  
2 macrophages induce the expression of the macrophage-specific receptor CD163 in malignant cells.  
3 *Int J Cancer*. 2012;131(10):2320-31. doi:10.1002/ijc.27506.
- 4 50. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S et al. Reducing waste from  
5 incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267-76.  
6 doi:10.1016/S0140-6736(13)62228-X.
- 7 51. Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PM, Korevaar DA et al. Increasing value and  
8 reducing waste in biomedical research: who's listening? *Lancet*. 2016;387(10027):1573-86.  
9 doi:10.1016/S0140-6736(15)00307-4.
- 10 52. Jankova L, Dent OF, Molloy MP, Chan C, Chapuis PH, Howell VM et al. Reporting in studies of  
11 protein biomarkers of prognosis in colorectal cancer in relation to the REMARK guidelines.  
12 *Proteomics Clin Appl*. 2015;9(11-12):1078-8. doi:10.1002/prca.201400177.
- 13 53. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ et al. CONSORT 2010  
14 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*.  
15 2010;340:c869. doi:10.1136/bmj.c869.
- 16 54. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement  
17 impact the completeness of reporting of randomised controlled trials published in medical journals?  
18 A Cochrane review. *Syst Rev*. 2012;1:60. doi:10.1186/2046-4053-1-60.
- 19 55. Coppus SF, van der Veen F, Bossuyt PM, Mol BW. Quality of reporting of test accuracy studies in  
20 reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD)  
21 initiative. *Fertil Steril*. 2006;86(5):1321-9. doi:10.1016/j.fertnstert.2006.03.050.
- 22 56. Smidt N, Rutjes AW, van der Windt DA, Ostelo RW, Bossuyt PM, Reitsma JB et al. The quality of  
23 diagnostic accuracy studies since the STARD statement: has it improved? *Neurology*. 2006;67(5):792-  
24 7. doi:10.1212/01.wnl.0000238386.41398.30.

- 1 57. Selman TJ, Morris RK, Zamora J, Khan KS. The quality of reporting of primary test accuracy studies  
2 in obstetrics and gynaecology: application of the STARD criteria. *BMC Womens Health*. 2011;11:8.  
3 doi:10.1186/1472-6874-11-8.
- 4 58. Bastuji-Garin S, Sbidian E, Gaudy-Marqueste C, Ferrat E, Roujeau JC, Richard MA et al. Impact of  
5 STROBE statement publication on quality of observational study reporting: interrupted time series  
6 versus before-after analysis. *PLoS One*. 2013;8(8):e64733. doi:10.1371/journal.pone.0064733.
- 7 59. da Costa BR, Cevallos M, Altman DG, Rutjes AW, Egger M. Uses and misuses of the STROBE  
8 statement: bibliographic study. *BMJ Open*. 2011;1(1):e000048. doi:10.1136/bmjopen-2010-000048.
- 9 60. Vandembroucke JP. STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION,  
10 COREQ, QUOROM, REMARK... and CONSORT: for whom does the guideline toll? *J Clin Epidemiol*.  
11 2009;62(6):594-6. doi:10.1016/j.jclinepi.2008.12.003.
- 12 61. Stevens A, Shamseer L, Weinstein E, Yazdi F, Turner L, Thielman J et al. Relation of completeness  
13 of reporting of health research to journals' endorsement of reporting guidelines: systematic review.  
14 *BMJ*. 2014;348:g3804. doi:10.1136/bmj.g3804.
- 15 62. Wang X, Chen Y, Yang N, Deng W, Wang Q, Li N et al. Methodology and reporting quality of  
16 reporting guidelines: systematic review. *BMC Med Res Methodol*. 2015;15:74. doi:10.1186/s12874-  
17 015-0069-z.
- 18 63. Marchionni L, Wilson RF, Wolff AC, Marinopoulos S, Parmigiani G, Bass EB et al. Systematic  
19 review: gene expression profiling assays in early-stage breast cancer. *Ann Intern Med*.  
20 2008;148(5):358-69.
- 21 64. Diamandis EP. The failure of protein cancer biomarkers to reach the clinic: why, and what can be  
22 done to address the problem? *BMC Med*. 2012;10:87. doi:10.1186/1741-7015-10-87.
- 23 65. Kleinert S, Horton R. How should medical science change? *Lancet*. 2014;383(9913):197-8.  
24 doi:10.1016/S0140-6736(13)62678-1.
- 25 66. Altman DG. Making research articles fit for purpose: structured reporting of key methods and  
26 findings. *Trials*. 2015;16:53. doi:10.1186/s13063-015-0575-7.

- 1 67. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A et al. Prognosis research strategy  
2 (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*. 2013;346:e5595.  
3 doi:10.1136/bmj.e5595.
- 4 68. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S et al. Prognosis  
5 Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381.  
6 doi:10.1371/journal.pmed.1001381.
- 7 69. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW et al. Prognosis research  
8 strategy (PROGRESS) 4: stratified medicine research. *BMJ*. 2013;346:e5793. doi:10.1136/bmj.e5793.
- 9 70. PROGnosis RESearch Strategy (PROGRESS) group. <http://progress-partnership.org/>.
- 10 71. STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative.  
11 <http://www.stratos-initiative.org/>.
- 12 72. Sauerbrei W, Abrahamowicz M, Altman DG, le Cessie S, Carpenter J, initiative S. STRengthening  
13 analytical thinking for observational studies: the STRATOS initiative. *Stat Med*. 2014;33(30):5413-32.  
14 doi:10.1002/sim.6265.
- 15 73. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JP. Reproducible Research Practices and  
16 Transparency across the Biomedical Literature. *PLoS Biol*. 2016;14(1):e1002333.  
17 doi:10.1371/journal.pbio.1002333.
- 18 74. Sauerbrei W, Hollander N, Riley RD, Altman DG. Evidence-based assessment and application of  
19 prognostic markers: The long way from single studies to meta-analysis. *Commun Stat-Theor M*.  
20 2006;35(7):1333-42. doi:10.1080/03610920600629666.
- 21 75. Andre F, McShane LM, Michiels S, Ransohoff DF, Altman DG, Reis JS et al. Biomarker studies: a  
22 call for a comprehensive biomarker study registry. *Nat Rev Clin Oncol*. 2011;8(3):171-6.  
23 doi:10.1038/nrclinonc.2011.4.
- 24 76. Altman DG. The time has come to register diagnostic and prognostic research. *Clin Chem*.  
25 2014;60(4):580-2. doi:10.1373/clinchem.2013.220335.

26

# 1 **Supporting Information**

2

3 **S1 Doc: REMARK checklist**

4 **S2 Doc: Eligibility criteria for selection of studies**

5 **S3 Doc: References of selected studies**

6 **S4 Doc: Data extraction form**

7 **S5 Doc: Assessed and discarded items of REMARK checklist – Reasons and definitions**

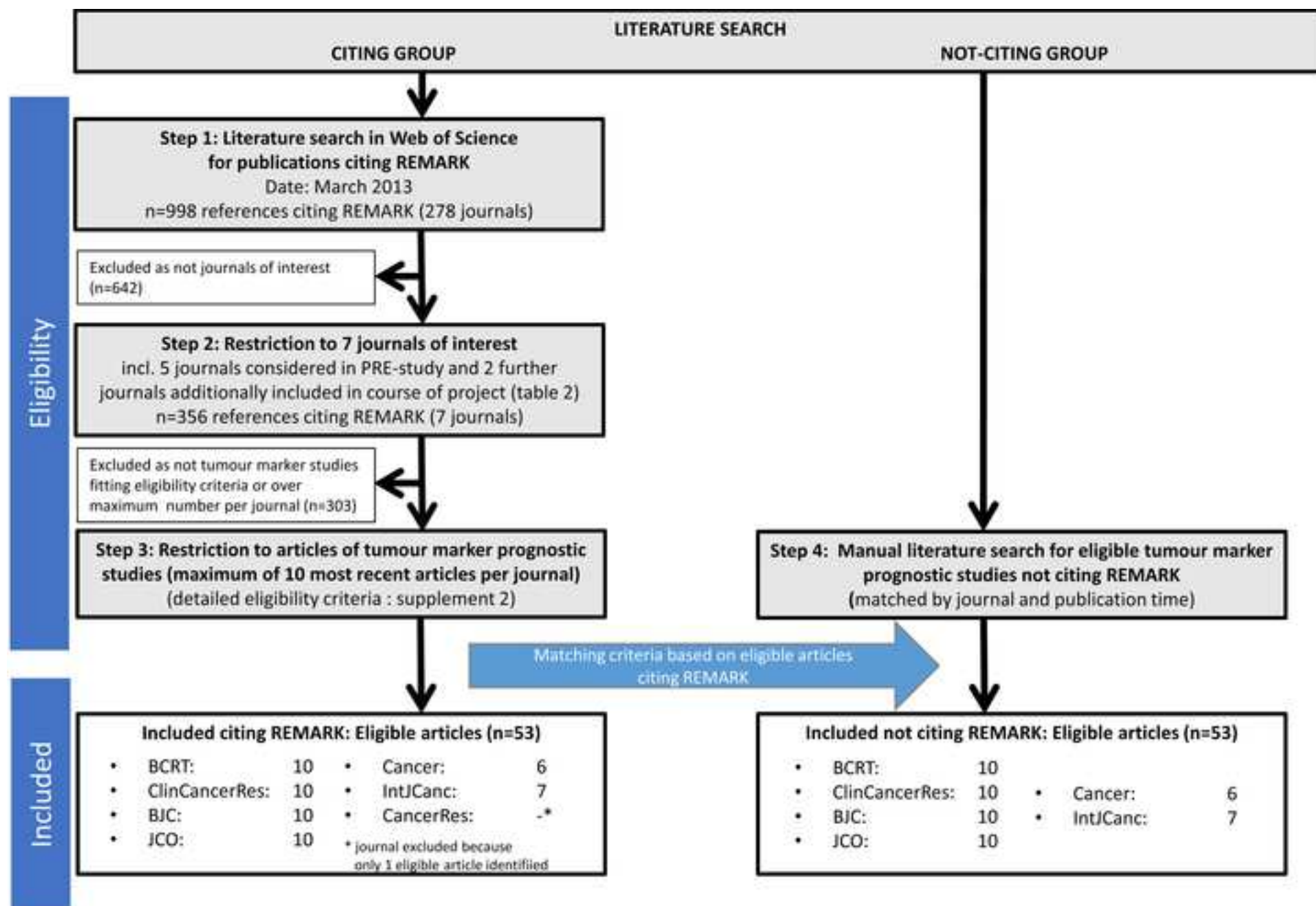
8

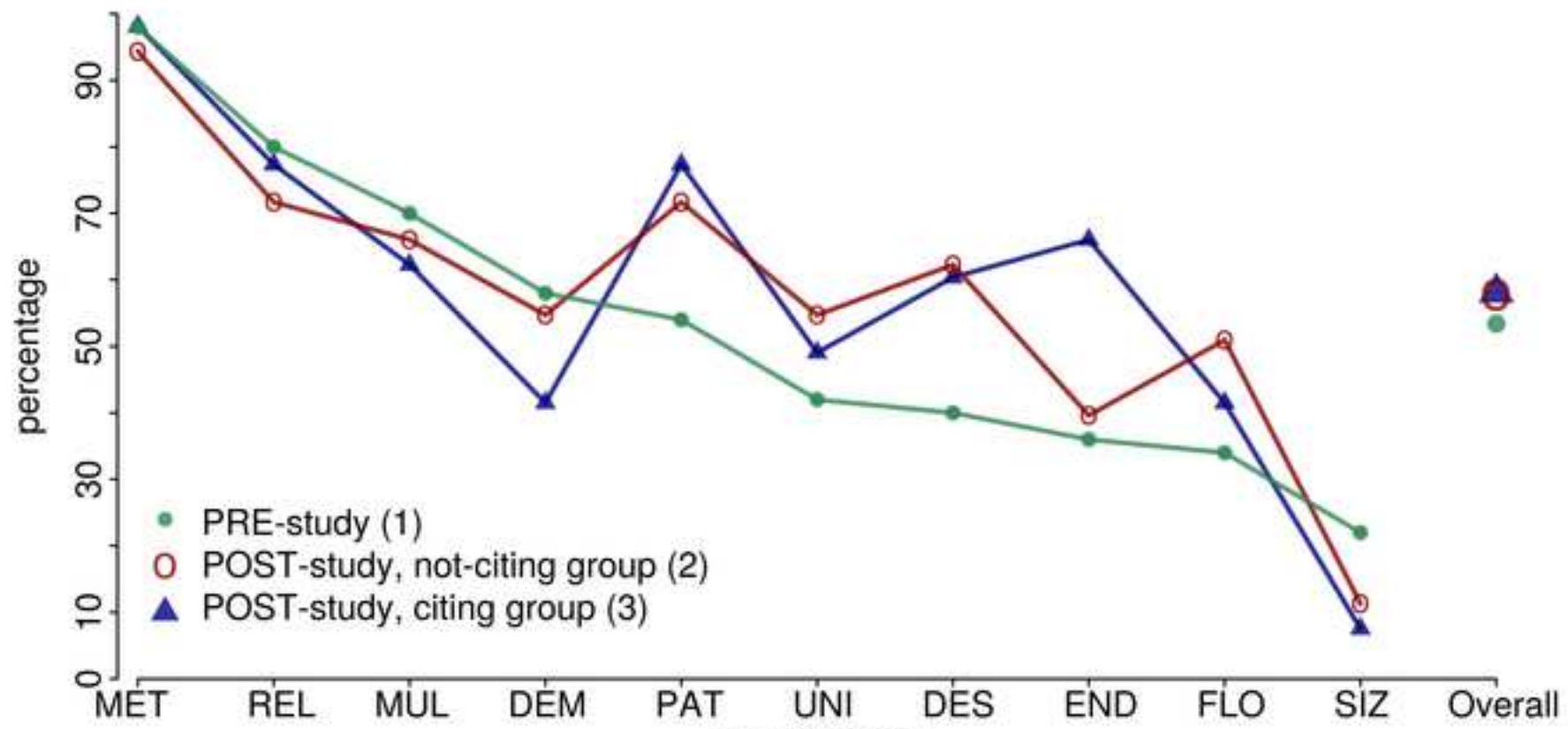
9 **S1 Table: Selected articles over time**

10 **S2 Table: Summary statistics**

11

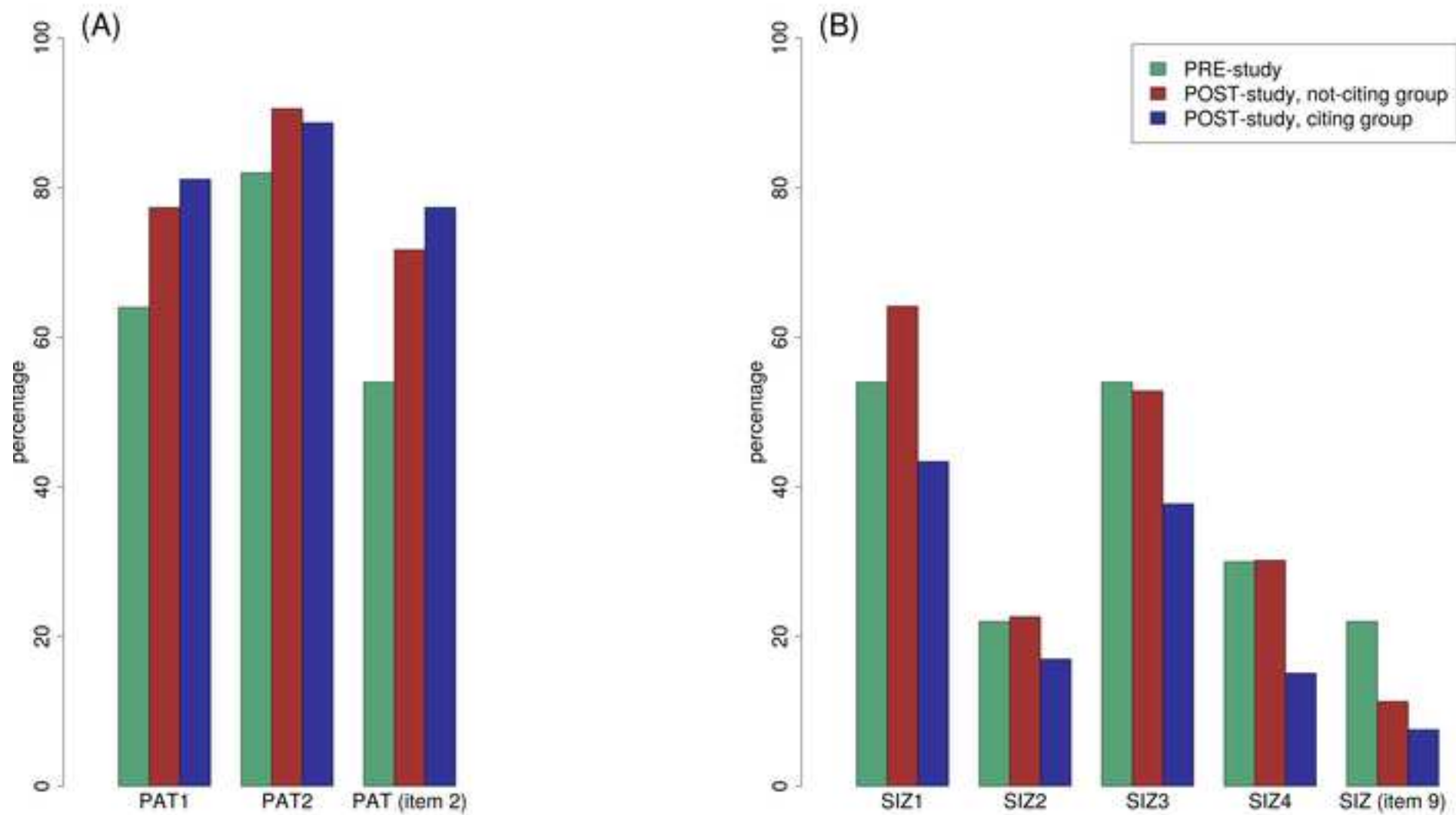
12 **S1 Data: Analysed data** (*additional excel-file*)





	MET	REL	MUL	DEM	PAT	UNI	DES	END	FLO	SIZ	Overall
(1)	98	80	70	58	54	42	40	36	34	22	53.4
(2)	94	72	66	55	72	55	62	40	51	11	57.7
(3)	98	77	62	42	77	49	60	66	42	8	58.1







Click here to access/download  
**Supporting Information**  
Sx\_Docs\_Tables.pdf



Click here to access/download  
**Supporting Information**  
S1\_Data.xlsx



Click here to access/download

**Other**

ReportingChecklist.pdf



1 **Did the reporting of prognostic studies of tumour markers**

2 **improve since the introduction of REMARK guideline?**

3 **A comparison of reporting in published articles.**

4 **Short title:** Evaluation of reporting quality of prognostic studies of tumour markers

5

6 Peggy Sekula<sup>1\*</sup>, Susan Mallett<sup>2</sup>, Douglas G Altman<sup>3</sup>, Willi Sauerbrei<sup>1</sup>

7

8 1 Institute for Medical Biometry and Statistics, Faculty of Medicine and Medical Center – University of Freiburg,

9 Freiburg, Germany

10 2 Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham, UK

11 3 Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal

12 Sciences, University of Oxford, Oxford, United Kingdom

13

14 \* Corresponding author

15 E-Mail: ps@imbi.uni-freiburg.de (PS)

# 1 **Abstract**

2 Although biomarkers are perceived as highly relevant for future clinical practice, few biomarkers  
3 reach clinical utility for several reasons. Among them, poor reporting of studies is one of the major  
4 problems. To aid improvement, reporting guidelines like REMARK for tumour marker prognostic  
5 (TMP) studies were introduced several years ago. The aims of this project were to assess whether  
6 reporting quality of TMP-studies improved in comparison to a previously conducted study assessing  
7 reporting quality of TMP-studies (PRE-study) and to assess whether articles citing REMARK (citing  
8 group) are better reported, in comparison to articles not citing REMARK (not-citing group).

9 For the POST-study, recent articles citing and not citing REMARK (53 each) were identified in selected  
10 journals through systematic literature search and evaluated in same way as in the PRE-study. Ten of  
11 the 20 items of the REMARK checklist were evaluated and used to define an overall score of  
12 reporting quality.

13 The observed overall scores were 53.4% (range: 10%-90%) for the PRE-study, 57.7% (range: 20%-  
14 100%) for the not-citing group and 58.1% (range: 30%-100%) for the citing group of the POST-study.

15 While there is no difference between the two groups of the POST-study, the POST-study shows a  
16 slight but not relevant improvement in reporting relative to the PRE-study. Not all the articles of the  
17 citing group, cited REMARK appropriately. Irrespective of whether REMARK was cited, the overall  
18 score was slightly higher for articles published in journals requesting adherence to REMARK than for  
19 those published in journals not requesting it: 59.9% versus 51.9%, respectively.

20 Several years after the introduction of REMARK, many key items of TMP-studies are still very poorly  
21 reported. A combined effort is needed from authors, editors, reviewers and methodologists to  
22 improve the current situation. Good reporting is not just nice to have but is essential for any research  
23 to be useful.

# 1 Introduction

2 Major advances in molecular biology and in analytical laboratory methods including new (high-  
3 throughput) technologies have enabled the detection and the measurement of a wide range of  
4 biomarkers in the human body. This has led to an increasing number of studies assessing the utility of  
5 biomarkers in a medical context [1, 2]. In this regard, a biomarker is an objectively measured  
6 characteristic with biological, clinical, genetic, histological or pathological background [3]. Biomarker  
7 measurement can be based on a single assessment or on a combination of information from several  
8 assessments (e.g. scores) [1, 4]. Biomarkers are already used successfully and routinely in different  
9 areas of medicine (e.g. serum creatinine to assess kidney function [5, 6]) and are perceived as highly  
10 relevant for future clinical practice using stratified or personalized medicine, where biomarkers may  
11 be useful to assist medical decision making, ideally underpinned by recommendations in clinical  
12 guidelines. Areas of biomarker use include but are not restricted to different aspects of patient care  
13 [2, 4, 7, 8]:

- 14 • screening of people to allow early detection of diseases,
- 15 • differential diagnosis of patients,
- 16 • stratification of patients for treatments,
- 17 • monitoring of treatment response and treatment compliance, and
- 18 • identification of risk groups related to patients' prognosis.

19 Biomarkers are also useful in the discovery and development of new treatments, through their role  
20 in elucidation of disease processes [4, 8]. Additionally, biomarkers are important in the design of  
21 studies and trials, allowing stratification of participants and use as surrogate endpoints [8, 9].

22 There are several important steps to establish the clinical value of a particular biomarker,  
23 including well-designed and well-reported clinical studies [1, 9-11]. Yet very few biomarkers have  
24 established clinical value [1, 12, 13], as exemplified by cancer research where it is estimated that  
25 fewer than 1% of biomarkers originally proposed as important have entered clinical practice [14].  
26 Researchers have investigated reasons for this unsatisfactory situation. Different types of failures

1 were distinguished, especially where results of subsequent studies contradict preceding study results  
2 [14-16]. There are major concerns that the poor quality of studies can lead to misleading results and  
3 consequently mistaken claims of utility [1, 3, 12, 17, 18].

4         As biomarker studies can be challenging, methodologists have highlighted the need for more  
5 transparency, standardization and harmonization to improve studies [11, 13, 19-23]. Overall, this will  
6 not only help to improve quality of individual studies but also enhance the ability to compare results  
7 between studies – a prerequisite for evidence synthesis and meta-analysis. A typical example is  
8 provided by p53. Since the early 1990s, p53 has been measured by immunohistochemistry and  
9 assessed as a potential prognostic biomarker in bladder cancer in many studies. Although  
10 researchers invested a lot of effort, time and money, the research question is still unanswered [24-  
11 28]. This situation is a consequence of many different methodological issues, such as small study  
12 populations and variation in study methods resulting in differences in the handling of measurements  
13 (e.g. different cutpoints used to define positive biomarker results).

14         Poor reporting is another major problem in these studies. Many biomarker studies are never  
15 reported at all and there is evidence that publication is linked to study results; Kyzas *et al* found that  
16 <1.5% of published prognostic marker studies were found to have only “negative” results [29].  
17 Within published studies, there are problems with the selective reporting of results and with the  
18 poor quality of reporting of methods and results [3, 21]. For tumour marker prognostic studies (TMP-  
19 studies), evidence for bad reporting has been provided [30]. In general, publications that are of poor  
20 quality can be essentially considered as a waste of research resources [31]. Worse, poor reporting in  
21 published studies might lead to incorrect conclusions about the evidence relating to a specific  
22 question.

23         To help overcome issues regarding the poor quality of reporting, guidelines for specific  
24 research areas were introduced. A valuable research hub is provided by the EQUATOR Network  
25 providing searchable access to reporting guidelines appropriate for many study designs and specific  
26 study features [32]. Among others, the REMARK guideline (short: REMARK) is a reporting guideline



1 specifically for TMP-studies assessing biomarkers in relation to future health outcomes in cancer  
2 patients. This guideline was published in seven journals in 2005/6 [33-39]. For convenience, the  
3 authors provided a checklist of 20 items addressing different parts of a manuscript. REMARK can be  
4 used by authors, editors and reviewers to check the reporting quality of a study report (**S1 Doc**). In  
5 addition, an extensive 'Explanation and Elaboration' (E&E) article was published in 2012, providing  
6 detailed information and examples of good reporting practice for each of these checklist items [40,  
7 41]. The need for REMARK was supported by a study that strikingly showed the poor reporting  
8 quality of 50 TMP-studies published in 2006-7 [30]. Because of the usual delay before an article is  
9 published, it is most unlikely that the authors of the assessed articles knew REMARK at the time of  
10 writing their manuscript (pre-REMARK period).

11 The aim of this project was to evaluate whether the quality of reporting of such studies has  
12 improved since the publication of REMARK (post-REMARK period). We repeated the previous study  
13 (short: PRE-study) using articles published between 2007 to 2012 (short: POST-study) using methods  
14 and definitions as similar as possible, to allow a fair comparison with previous findings [30]. Some  
15 TMP-studies cite the REMARK guidelines demonstrating awareness of REMARK, sometimes because  
16 journals like *Breast Cancer Research and Treatment (BCRT)* ask for adherence to REMARK in  
17 submitted manuscripts [42]. In contrast, authors of articles not citing the guideline are more likely to  
18 be unaware of REMARK or may not be using the checklist. In this study, we also addressed the  
19 question whether citing the REMARK guideline or not is related to the reporting quality. In summary,  
20 the two aims of the project are:

- 21 1. Has there been any improvement in reporting quality since introduction of REMARK?
- 22 2. Is reporting better in studies citing REMARK?

23

24

# 1 **Material and Methods**

2 Because only published data from studies in humans were utilized, no approval from an ethic  
3 committee was obtained.

4 To allow a direct comparison to the previous work [30], the POST-study was designed in a  
5 very similar way (choice of journals, study selection, data extraction). In this study, two groups of  
6 publications were distinguished: (A) publications that cited one of the seven REMARK publications  
7 (citing group) and (B) publications that did not cite REMARK (not-citing group) [33-39]. Similarly to  
8 the PRE-study, it was planned to include 50 articles per group as sufficient size in this methodological  
9 study to address questions of interest.

10

## 11 **1 Literature search**

12 To identify TMP-studies citing REMARK, a literature search was done in Web of Science in March  
13 2013. References of all publications citing at least one of the REMARK publications were extracted  
14 and imported into Endnote [33-39]. After removal of duplicates (n=72), 998 articles published in 278  
15 different journals were identified. Among them, 134 publications were in one of the five previously  
16 considered cancer journals: *Cancer [Canc]*, *Cancer Research [CaRes]*, *International Journal of Cancer*  
17 *[IJC]*, *Journal of Clinical Oncology [JCO]*, *Clinical Cancer Research [CCR]*.

18 The 134 identified articles published in the five journals considered in the PRE-study were  
19 then examined to identify for each journal the 10 most recent TMP-studies that cited REMARK. A  
20 detailed description of the eligibility criteria can be found in the **S2 Doc**. Essentially, studies assessing  
21 the prognostic impact of a specific biomarker on an outcome of clinical importance (e.g. cancer-  
22 specific survival) in cancer patients were eligible. The search revealed 10 articles each from *JCO* and  
23 *CCR*, 7 from *IJC*, 6 from *Canc* and 1 from *CaRes*. Because of this result, we decided to exclude *CaRes*  
24 from further consideration and to include two further cancer journals (*Breast Cancer Research and*  
25 *Treatment [BCRT]* and *British Journal of Cancer [BJC]*) for which 10 articles each could be identified.

1 Altogether, the citing group comprised 53 articles. Although about 80% of the included manuscripts  
2 were published in 2011 and 2012, a few dated back to 2007.

3 To identify publications not citing REMARK, we aimed to obtain for each article citing  
4 REMARK another article from the same journal that did not, closely matched in time (publication  
5 year and, if possible, issue). The same number of articles (n=53) was identified forming the not-citing  
6 group.

7 The described search is depicted in **Fig 1**. The references of all selected articles are listed in  
8 **S3 Doc**.

9

10 **Fig 1: Literature search – flow chart**

11

## 12 **2 Data extraction**

13 For all 106 articles from the six journals we obtained the full text. For data extraction, we used the  
14 same standardized form that had been used in the PRE-study (**S4 Doc**) [30]. This form lists several  
15 elements (specific questions) addressing different items of the REMARK checklist. The focus of data  
16 extraction led on information related to methods and results of a study. Because of the general  
17 character of each checklist item, a specific item is often described by more than 1 element of the  
18 data extraction form.

19 To ensure good comparability of extracted data with past results, a pilot data extraction for  
20 eight articles was done in duplicate by the author (SM) who mainly did the data extraction in the  
21 PRE-study and another author (PS) who was responsible for it in the POST-study. Results of these  
22 extractions were compared and differences clarified before data extractions were done for the  
23 remaining articles by PS alone.

24 For articles in which several biomarkers were assessed in a study in parallel, the biomarker  
25 first mentioned in title or abstract for which a multivariable analysis was done was defined to be the  
26 focus of the data extraction. A similar approach was used when different study populations were

1 assessed within a single article. Two groups of time-to-event outcomes were distinguished: death-  
2 related outcomes (overall survival, cancer-specific survival) and other time-to-event outcomes  
3 (disease-free survival, time until recurrence/relapse). Similarly, when several outcomes were  
4 assessed in a study the data extraction focused on the outcome that was first mentioned in title or  
5 abstract for which a multivariable analysis had been conducted.

6 Importantly, this project focuses only on the assessment of reporting quality and not on the  
7 general appropriateness of methods, including study design, assessed biomarkers, statistical  
8 methods and outcomes considered.

9

### 10 **3 Analyses**

11 We addressed our first aim on the improvement over time by comparing the results obtained in the  
12 PRE-study to those of the not-citing group. The second aim on difference in reporting when citing or  
13 not citing REMARK was addressed by comparing the results for the citing and not-citing groups within  
14 the current POST-study.

15 The intended comparisons were descriptively conducted with respect to 10 of the 20 items of  
16 the REMARK checklist that are related to methods and results of a manuscript (**Table 1**). For each  
17 article, we evaluated whether information for each item was provided (yes/no) by combining  
18 extracted information of elements assigned to that item. Details regarding selection of checklist  
19 items and definitions how items were evaluated are provided in **S5 Doc**. Finally, for each article an  
20 overall score was obtained as the percentage of items addressed out of 10.

21

22

1 **Table 1: Overview of the 10 assessed items of the REMARK checklist**

No.	Manuscript part	Item of REMARK checklist	Short description	Abbreviation used in article
1	Methods	2	<b>PAT</b> ient characteristics	PAT
2		6	Study <b>DES</b> ign: patient selection & time period	DES
3		7	Clinical <b>END</b> points	END
4		9	Rationale for sample <b>SIZ</b> e	SIZ
5		10	All statistical <b>METH</b> ods	MET
6	Results	12	<b>FLO</b> w of patients	FLO
7		13	Distribution of <b>DEM</b> ographic characteristics	DEM
8		14	<b>REL</b> ationship between marker and standard variables	REL
9		15	<b>UNI</b> variate analyses	UNI
10		16	<b>MULT</b> ivariable analyses	MUL

2

3           Only 10 checklist items were included in the assessment of adherence to REMARK as we used  
4 only items we could assess objectively and that could be assessed on TMP-studies from any research  
5 area. Items 1, 19 and 20 referring to the introduction and the discussion of an article were  
6 considered too subjective and require subject-specific expert knowledge, and so had not been  
7 included in the data extraction form that was already used in the PRE-study. Similarly, the seven  
8 items 3, 4, 5, 8, 11, 17 and 18 referring to the methods and the results of an article were excluded  
9 because their evaluation essentially requires profound expert knowledge with respect to the medical  
10 background and methodology. For more details, see **S5 Doc**.

11

12

## 1 **4 Reporting**

2 This study assesses the reporting quality of published TMP-studies. For such a ‘research on research’-  
3 project, no specific reporting guideline is available. The current project, however, shows some  
4 features (observational kind, literature search) that allow us to use different reporting statements as  
5 guidance. Specifically, we used the STROBE guideline for general aspects of the project and the  
6 PRISMA statement for aspects around literature search [43, 44].

7

8

# 1 Results

2

## 3 1 Selected journals and assessed articles

### 4 Overall

5 The POST-study was planned to use the same journals as much as was feasible, but some changes to  
 6 journals was required for practical reasons (**Table 2**). All journals are of higher impact (impact factor  
 7 2012 >4). Three journals (*BCRT*, *BJC*, *JCO*) belong to the group of journals that have published  
 8 REMARK. These three journals and *CCR* explicitly ask authors submitting a manuscript for adherence  
 9 to REMARK (**Table 2**).

10

11 **Table 2: Cancer journals included in the PRE-study and in the POST-study**

Journal (alphabetical order)	PRE-study		POST-study		Impact factor <sup>†</sup> 2012	Publication of REMARK	Author instructed to adhere to REMARK	
		N assessed articles		N assessed articles <sup>*</sup>			02/2009 <sup>‡</sup>	08/2014
<i>BCRT</i>	-	-	✓	10/10	4.5	YES [38]	YES [42]	YES
<i>BJC</i>	-	-	✓	10/10	5.1	YES [33]	UNK	YES
<i>Canc</i>	✓	10	✓	6/6	5.2	NO	NO	NO
<i>CaRes</i>	✓	10	-	- <sup>§</sup>	8.6	NO	NO	NO
<i>CCR</i>	✓	10	✓	10/10	7.8	NO	YES	YES
<i>IJC</i>	✓	10	✓	7/7	6.2	NO	NO	NO
<i>JCO</i>	✓	10	✓	10/10	18.0	YES [35]	YES	YES

12 <sup>\*</sup> N citing group/n not-citing group; <sup>†</sup> source: InCites™ Journal Citation Reports; <sup>‡</sup> check was done within the PRE-study; <sup>§</sup> journal was

13 excluded because only one eligible article citing REMARK was identified; UNK=unknown

14

1           The sample included 53 articles in both the citing group and the not-citing group (total  
2 n=106). Similarly to the PRE-study, the distribution of cancer sites was diverse. As a consequence of  
3 the additional inclusion of *BCRT*, however, the proportion of breast and/or ovarian cancer studies  
4 was higher in the current sample (PRE-study: 30%, POST-study: 44%). Articles in the not-citing and  
5 citing groups were well matched by journal, year and issue (**S1 Table**).

6

## 7 **Citing group**

8           At least one of the REMARK publications was referenced in all the articles assigned to the citing  
9 group. Since REMARK is a methodological tool, its citation is expected to be given in the methods  
10 section of the article, with a statement like “The study is reported in accordance to the REMARK  
11 guideline”. Although REMARK was indeed cited most often in the methods section (n=39, 74%), some  
12 citations appeared in other sections of the manuscripts. The statements in which REMARK was cited  
13 varied greatly. While some authors correctly referred to the reporting of the study, other authors  
14 referred to REMARK in relation to the design, the conduct and the analysis of the study. For example,  
15 the statement “*This analysis was conducted according to the reporting recommendations for tumor  
16 marker guidelines for prognostic studies ...*” was provided by the authors in the methods section [45].  
17 Other statements are difficult to understand, such as “*Protein expression was evaluated using a  
18 semiquantitative weighted histoscore method by two observers as previously described ... in  
19 accordance with the Reporting Recommendations for Tumour Marker Prognostic Studies (REMARK)  
20 guidelines ...*” [46].

21           Three manuscripts were accompanied by a completed REMARK checklist [47-49]; two of  
22 these had overlapping authorship [47, 48]. For unknown reasons, none of these lists cover the full  
23 REMARK checklist of 20 items. Moreover, some explanations were difficult to assess, for example  
24 regarding the item ‘Flow of patients’, the authors stated “*This is not a staged analysis. The evaluated  
25 cohort is described ...*” [48].

26



## 1 **2 Comparison of reporting quality**

### 2 **Not-citing group (POST-study) versus PRE-study**

3 Overall, there was a slight but not relevant improvement in the mean overall score: PRE-study 53.4%  
4 (range: 10%-90%), not-citing group of POST-study 57.7% (range: 20%-100%, Wilcoxon rank sum test:  
5  $p = 0.33$ , **Fig 2**). This small difference, however, vanished when we included only articles published in  
6 the four journals assessed in both periods: PRE-study 56.5% (range: 10%-80%,  $n=40$ ), not-citing group  
7 of POST-study 56.4% (range: 20%-80%,  $n=33$ ). Some items showed an improvement in reporting from  
8 the past to the present, while others showed a decline (**S2 Table**). An improvement, for example, was  
9 visible for item 2 'Patient characteristics' (PAT): PRE-study 54%, POST-study 72%. In this case, the  
10 improvement was also visible for the single assigned elements like the element 'Selection of patients'  
11 (PAT1) showing improvement from 64% in the PRE-study to 77% in the not-citing group (**Fig 3A**). In  
12 contrast, a remarkable decline from past to present was seen for item 9 'Rationale for sample size'  
13 (SIZ; **Fig 3B, S5 Doc**). Overall, there remains much room for improvement of reporting.

14

15 **Fig 2: Percentages of articles adequately reporting information for 10 selected items of the**  
16 **REMARK checklist.**

17 The items are sorted by percentages obtained in the PRE-study [30]. See Table 1 or S2 Table for  
18 explanation of abbreviations used for different checklist items.

19

20 **Fig 3: Percentages of articles adequately reporting information for two checklist items and their**  
21 **single elements respectively assigned.**

22 (A) Checklist item 2 'Patient characteristics' (PAT), (B) checklist item 9 'Rationale for sample size'  
23 (SIZ). See S2 Table for explanation of abbreviations used for different elements of data extraction and  
24 checklist items.

25

1 **Citing group versus not-citing group (POST-study)**

2 When comparing the not-citing group and the citing group, there was essentially no difference in  
3 mean scores: not-citing group 57.7% (range: 20%-100%), citing group 58.1% (range: 30%-100%, **Fig**  
4 **2**). Again, some single checklist items showed an improvement in reporting from the past to the  
5 present, while others showed a decline. Most pronounced, item 7 ‘Clinical endpoints’ (END) was  
6 reported better in the citing group than in the non-citing group (40% vs 66%, respectively; **S2 Table**),  
7 whereas it was the other way around for item 13 ‘Distribution of demographic characteristics’ (DEM;  
8 55% vs 42%). **Fig 3** similarly illustrates observed percentages for item 2 (PAT) and item 9 (SIZ).

9

10 **Additional analysis**

11 Because we observed some unexpected statements by authors citing REMARK which could imply a  
12 lack of understanding of REMARK as a reporting guideline, an additional comparison was made of  
13 articles grouped by journals requesting (4 journals, 80 articles) or not requesting (2 journals, 26  
14 articles) adherence to REMARK, irrespective whether authors cited or not cited REMARK (**Table 2**).  
15 On average, the overall score for articles published in journals requesting adherence to REMARK was  
16 higher (59.9%) than for the other group (51.9%). This ordering was also present for each single  
17 checklist item.

18

19

# 1 Discussion

2 Several years after REMARK was introduced, and with many published discussions of the reporting  
3 quality in health research and prominence given to the role of poor reporting in contributing to  
4 research waste, some improvement of the reporting quality of TMP-studies was expected [30, 50,  
5 51]. However, our assessment of articles from the post-REMARK period did not reveal any relevant  
6 improvement over the quality of articles assessed in the earlier study [30]. The overall reporting  
7 quality is still very poor. Authors still frequently fail to report important aspects of their study such as  
8 the source of the study population, fully defined clinical endpoints, and an explanation of the sample  
9 size.

10 Moreover, we observed essentially no difference in reporting quality when comparing  
11 articles citing and not citing REMARK. Because citing REMARK means the author of the respective  
12 article is aware of the guideline, one would expect to see superior reporting quality compared to  
13 articles not citing REMARK. Our findings, however, raise the question of whether the main scope of  
14 REMARK is really understood. To overcome any misunderstanding the REMARK group already  
15 published a manuscript that elaborates and explains each item of the REMARK checklist in detail [40,  
16 41]. However, authors of articles assessed in this project (published  $\leq 2012$ ) could not have known  
17 this amendment because it was published in 2012.

18 Because of these disappointing results we decided to conduct an additional unplanned  
19 comparison between reporting qualities of articles published in journals requesting or not requesting  
20 adherence to REMARK in the submission guidelines. This revealed somewhat better reporting in the  
21 group of articles published in journals requesting adherence to REMARK.

22

23

## 1 **1 Limitations of study**

2 To allow a fair comparison of results between past and current assessments, the current project was  
3 designed to be as similar as possible to the previous study. Also, the current team largely overlaps  
4 with the team of the past study. Furthermore, all the documents including the data extraction form  
5 could be utilized. A pilot study was conducted to ensure comparability between data extractions in  
6 the past and current projects. Still, some systematic differences between the two surveys cannot be  
7 ruled out. In addition although judged as sufficient to address the methodological research question,  
8 the number of studies assessed was relatively small in both the pre-study and the current study.

9 One obvious limitation of this study is that we could not identify enough articles in all  
10 journals considered in the first study, so two new journals (*BCRT, BJC*) were included. Since both  
11 additional journals published REMARK and requesting adherence to it, the overall result might be  
12 biased. For this reason, an additional analysis was conducted by restricting articles to those published  
13 in the four journals *Canc, CCR, IJC* and *JCO* that were considered in both assessments. As result, the  
14 small improvement observed in the overall sample vanished. Overall we found no improvement in  
15 reporting quality of prognostic factor studies in the first (about six) years since REMARK was  
16 published. Repeating the investigation with papers published after more than ten years (say in 2016)  
17 may provide better results.

18 Another issue relates to the overall score used to evaluate quality of reporting. The overall  
19 score included summation of sufficiently reported REMARK items, often based on several elements  
20 of the data extraction form. For transparency, a description of the overall score and detailed results  
21 are provided in the supporting information (**S5 Doc, S2 Table**).

22

23

## 1 **2 Our findings in the context of published literature**

2 To our knowledge, there is just one other published study assessing quality of TMP-studies, which  
3 reviews studies of prognostic markers for colorectal cancer published in 2009-11, a slightly earlier  
4 period to the current project [52]. The authors assessed adherence to the complete REMARK  
5 checklist and found a mean score of 60 out of 78, but still emphasize deficiencies in reporting similar  
6 to those seen in our study across all cancers.

7 Concern about reporting quality applies across all areas of health research. To overcome this  
8 problem reporting guidelines for many different study designs and research areas are available [32].  
9 Similarly to our project, other study groups also assessed the question of whether reporting quality  
10 improved over time. For randomized controlled trials and in relation to the CONSORT statement,  
11 modest improvement in reporting quality was reported but reporting was still considered suboptimal  
12 [53, 54]. For other guidelines like STARD or STROBE, some slight improvements were also reported  
13 [55-58]. The current study on REMARK is essentially in line with those other reported results.

14 Da Costa *et al* systematically examined reasons for citing STROBE guideline [43, 59]. Similar  
15 to our observations, the authors reported that the guideline is often used inappropriately. These  
16 observations raise doubts on the general understanding of reporting guidelines and their aim, as  
17 already discussed in 2008 [60].

18 Evidence of a relation between reporting quality and endorsement of reporting guidelines by  
19 journals is limited [54, 61]. Our data suggest that a request of adherence by the journal might be  
20 useful. In order to provide conclusive evidence, well-planned prospective studies in cooperation with  
21 editors are needed to explore and enhance journal editor led interventions to improve reporting  
22 [61].

23 Based on our experience in the current project, we became aware that expert knowledge of  
24 the research subject and methods is often required to evaluate details needed for good reporting.  
25 Editors and reviewers may find it hard to recruit and focus experts on reporting as well as results of  
26 research studies. For authors writing a manuscript, access to sufficient expertise should be easier

1 because the research team should include experts relevant to the clinical and methodological aspects  
2 of a study. On the other hand, reporting guidelines are misunderstood by many authors [62], and  
3 further initiatives like the E&E paper for REMARK may be very helpful [40, 41].

4

### 5 **3 Quality of medical research in general**

6 In general, the quality of medical research, including other aspects besides reporting, has been  
7 criticized heavily in the last years [1, 14, 19, 31, 63-66]. To overcome these issues, several important  
8 contributions as well as the introduction of reporting guidelines have been seen recently. For  
9 example, the PROGnosis RESearch Strategy (PROGRESS) group published a series of articles to  
10 provide a framework on different aspects in prognostic research [11, 67-70]. Also, the STRATOS  
11 (STRengthening Analytical Thinking for Observational Studies) initiative was founded recently that  
12 aims to derive guidance documents related to design and analysis of observational studies [71, 72].

13 Overall, the need for transparency in medical research still appears to lack widespread  
14 acceptance and research endeavour [21, 23, 73]. Researchers remain insufficiently aware of the need  
15 to make their research clear and understandable to other researchers, as well as practising  
16 physicians, patients and other stakeholders (e.g. pharmaceutical companies, funding agencies).  
17 Particularly in medical research it is important that studies can be repeated by other research groups,  
18 requiring transparency through good reporting of research methods and results.

19 Registration of all studies and data sharing [1, 23, 73-76] have been recommended to  
20 improve knowledge of ongoing and past research. In this context, good reporting is a main  
21 prerequisite. Even a well-conducted and well-analysed study that is poorly reported can be  
22 considered as waste of resources.

23

24

# 1 **Conclusions**

2 Tumour marker prognostic studies are still very poorly reported. To improve the situation the  
3 REMARK recommendations need to be followed, ~~which will require combined efforts of authors,~~  
4 ~~editors, reviewers and methodologists.~~ An extensive guidance document is available to facilitate the  
5 ~~understanding of REMARK.~~ However, this study is another example illustrating that publication of  
6 guidelines is insufficient and that more pressure on authors, reviewers and editors is needed to  
7 improve on this unfortunate situation. We support the proposal of one reviewer of this manuscript  
8 that an electronic checklist (a web-based form of the checklist on which the authors can indicate  
9 where in the manuscript information for an item is addressed) can be a valuable instrument of the  
10 submission process. Ideally, such an electronic document can also provide further information about  
11 the reporting items. We hope that more journals will be willing to request such checklists in their  
12 submission process. Good reporting is not just nice to have. It is essential for any research to be  
13 useful but also for the limitations of research to be understood. Good reporting is also essential for  
14 systematic reviews that bring together and overview research studies to achieve a high level of  
15 evidence.

# 1 **Acknowledgement**

2 We thank Saskia Motschall, Clemens Wachter and Corinna Roller (student helpers) for their support  
3 during the literature search, article acquisition and selection of published studies.

4 The article processing charge was funded by the German Research Foundation (DFG) and the  
5 University of Freiburg in the funding programme Open Access Publishing.

6



# 1   **References**

- 2   1. Riley RD, Sauerbrei W, Altman DG. Prognostic markers in cancer: the evolution of evidence from  
3   single studies to meta-analysis, and beyond. *Br J Cancer*. 2009;100(8):1219-29.  
4   doi:10.1038/sj.bjc.6604999.
- 5   2. Rifai N, Watson ID, Miller WG. Commercial immunoassays in biomarkers studies: researchers  
6   beware! *Clin Chem*. 2012;58(10):1387-8. doi:10.1373/clinchem.2012.192351.
- 7   3. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Jones DR, Heney D et al. Reporting of prognostic  
8   markers: current problems and development of guidelines for evidence-based practice in the future.  
9   *Br J Cancer*. 2003;88(8):1191-8. doi:10.1038/sj.bjc.6600886.
- 10   4. Anderson DC, Kodukula K. Biomarkers in pharmacology and drug discovery. *Biochem Pharmacol*.  
11   2014;87(1):172-88. doi:10.1016/j.bcp.2013.08.026.
- 12   5. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI et al. A new equation to  
13   estimate glomerular filtration rate. *Ann Intern Med*. 2009;150(9):604-12.
- 14   6. Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T et al. Estimating glomerular  
15   filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20-9.  
16   doi:10.1056/NEJMoa1114248.
- 17   7. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive  
18   biomarkers. *J Natl Cancer Inst*. 2009;101(21):1446-52. doi:10.1093/jnci/djp335.
- 19   8. Micheel CM, Nass SJ, Omenn GS. Evolution of Translational Omics: Lessons Learned and the Path  
20   Forward. National Academies Press. 2012. doi:DOI: 10.17226/13297.
- 21   9. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer*.  
22   1994;69(6):979-85.
- 23   10. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M et al. Phases of biomarker  
24   development for early detection of cancer. *J Natl Cancer Inst*. 2001;93(14):1054-61.

- 1 11. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA et al. Prognosis Research  
2 Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med.* 2013;10(2):e1001380.  
3 doi:10.1371/journal.pmed.1001380.
- 4 12. McShane LM, Altman DG, Sauerbrei W. Identification of clinically useful cancer prognostic  
5 factors: what are we missing? *J Natl Cancer Inst.* 2005;97(14):1023-5. doi:10.1093/jnci/dji193.
- 6 13. Plebani M. Lessons from controversy: biomarkers evaluation. *Clin Chem Lab Med.*  
7 2013;51(2):247-8. doi:10.1515/cclm-2013-0012.
- 8 14. Kern SE. Why your new cancer biomarker may never work: recurrent patterns and remarkable  
9 diversity in biomarker failures. *Cancer Res.* 2012;72(23):6097-101. doi:10.1158/0008-5472.CAN-12-  
10 3232.
- 11 15. Ioannidis JP. Evolution and translation of research findings: from bench to where? *PLoS Clin*  
12 *Trials.* 2006;1(7):e36. doi:10.1371/journal.pctr.0010036.
- 13 16. Ioannidis JP. Biomarker failures. *Clin Chem.* 2013;59(1):202-4.  
14 doi:10.1373/clinchem.2012.185801.
- 15 17. Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in  
16 breast cancer. *Breast Cancer Res Treat.* 1998;52(1-3):289-303.
- 17 18. Sauerbrei W. Prognostic factors. Confusion caused by bad quality design, analysis and reporting  
18 of many studies. *Adv Otorhinolaryngol.* 2005;62:184-200. doi:10.1159/000082508.
- 19 19. Altman DG. The scandal of poor medical research. *BMJ.* 1994;308(6924):283-4.
- 20 20. Ioannidis JP. Genetic and molecular epidemiology. *J Epidemiol Community Health.*  
21 2007;61(9):757-8. doi:10.1136/jech.2006.059055.
- 22 21. McShane LM, Hayes DF. Publication of tumor marker research results: the necessity for complete  
23 and transparent reporting. *J Clin Oncol.* 2012;30(34):4223-32. doi:10.1200/JCO.2012.42.6858.
- 24 22. Behrens T, Bonberg N, Casjens S, Pesch B, Bruning T. A practical guide to epidemiological practice  
25 and standards in the identification and validation of diagnostic markers using a bladder cancer  
26 example. *Biochim Biophys Acta.* 2014;1844(1 Pt A):145-55. doi:10.1016/j.bbapap.2013.07.018.

- 1 23. Peat G, Riley RD, Croft P, Morley KI, Kyzas PA, Moons KG et al. Improving the transparency of  
2 prognosis research: the role of reporting, data sharing, registration, and protocols. *PLoS Med*.  
3 2014;11(7):e1001671. doi:10.1371/journal.pmed.1001671.
- 4 24. Schmitz-Drager BJ, Goebell PJ, Ebert T, Fradet Y. p53 immunohistochemistry as a prognostic  
5 marker in bladder cancer. Playground for urology scientists? *Eur Urol*. 2000;38(6):691-9;discussion  
6 700. doi:20364.
- 7 25. Malats N, Bustos A, Nascimento CM, Fernandez F, Rivas M, Puente D et al. P53 as a prognostic  
8 marker for bladder cancer: a meta-analysis and review. *Lancet Oncol*. 2005;6(9):678-86.  
9 doi:10.1016/S1470-2045(05)70315-6.
- 10 26. Goebell PJ, Groshen SG, Schmitz-Drager BJ, International Study-Initiative on Bladder C. p53  
11 immunohistochemistry in bladder cancer--a new approach to an old question. *Urol Oncol*.  
12 2010;28(4):377-88. doi:10.1016/j.urolonc.2010.03.021.
- 13 27. Kamat AM, Hegarty PK, Gee JR, Clark PE, Svatek RS, Hegarty N et al. ICUD-EAU International  
14 Consultation on Bladder Cancer 2012: Screening, diagnosis, and molecular markers. *Eur Urol*.  
15 2013;63(1):4-15. doi:10.1016/j.eururo.2012.09.057.
- 16 28. Sanguedolce F, Bufo P, Carrieri G, Cormio L. Predictive markers in bladder cancer: do we have  
17 molecular markers ready for clinical use? *Crit Rev Clin Lab Sci*. 2014;51(5):291-304.  
18 doi:10.3109/10408363.2014.930412.
- 19 29. Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Almost all articles on cancer prognostic markers report  
20 statistically significant results. *Eur J Cancer*. 2007;43(17):2559-79. doi:10.1016/j.ejca.2007.08.030.
- 21 30. Mallett S, Timmer A, Sauerbrei W, Altman DG. Reporting of prognostic studies of tumour  
22 markers: a review of published articles in relation to REMARK guidelines. *Br J Cancer*.  
23 2010;102(1):173-80. doi:10.1038/sj.bjc.6605462.
- 24 31. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JP et al. Biomedical research:  
25 increasing value, reducing waste. *Lancet*. 2014;383(9912):101-4. doi:10.1016/S0140-6736(13)62329-  
26 6.

- 1 32. EQUATOR Network. <http://www.equator-network.org>.
- 2 33. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. REporting  
3 recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer. 2005;93(4):387-91.  
4 doi:10.1038/sj.bjc.6602678.
- 5 34. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. REporting  
6 recommendations for tumour MARKer prognostic studies (REMARK). Eur J Cancer. 2005;41(12):1690-  
7 6. doi:10.1016/j.ejca.2005.03.032.
- 8 35. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. Reporting  
9 recommendations for tumor marker prognostic studies. J Clin Oncol. 2005;23(36):9067-72.  
10 doi:10.1200/JCO.2004.01.0454.
- 11 36. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. Reporting  
12 recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst.  
13 2005;97(16):1180-4. doi:10.1093/jnci/dji237.
- 14 37. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. REporting  
15 recommendations for tumor MARKer prognostic studies (REMARK). Nat Clin Pract Oncol.  
16 2005;2(8):416-22.
- 17 38. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. REporting  
18 recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat.  
19 2006;100(2):229-35. doi:10.1007/s10549-006-9242-8.
- 20 39. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM et al. Reporting  
21 recommendations for tumor marker prognostic studies (remark). Exp Oncol. 2006;28(2):99-105.
- 22 40. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting recommendations for tumor marker  
23 prognostic studies (REMARK): explanation and elaboration. BMC Med. 2012;10:51.  
24 doi:10.1186/1741-7015-10-51.

- 1 41. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker  
2 Prognostic Studies (REMARK): explanation and elaboration. *PLoS Med.* 2012;9(5):e1001216.  
3 doi:10.1371/journal.pmed.1001216.
- 4 42. Hayes DF, Ethier S, Lippman ME. New guidelines for reporting of tumor marker studies in breast  
5 cancer research and treatment: REMARK. *Breast Cancer Res Treat.* 2006;100(2):237-8.  
6 doi:10.1007/s10549-006-9253-5.
- 7 43. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP et al. The  
8 Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement:  
9 guidelines for reporting observational studies. *Ann Intern Med.* 2007;147(8):573-7.
- 10 44. Moher D, Liberati A, Tetzlaff J, Altman DG, Prisma Group. Preferred reporting items for  
11 systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097.  
12 doi:10.1371/journal.pmed.1000097.
- 13 45. Debled M, MacGrogan G, Brouste V, Mathoulin-Pelissier S, Durand M, Mauriac L. Prognostic  
14 factors of early distant recurrence in hormone receptor-positive, postmenopausal breast cancer  
15 patients receiving adjuvant tamoxifen therapy: results of a retrospective analysis. *Cancer.*  
16 2007;109(11):2197-204. doi:10.1002/cncr.22667.
- 17 46. Spears M, Pederson HC, Lyttle N, Gray C, Quintayo MA, Brogan L et al. Expression of activated  
18 type I receptor tyrosine kinases in early breast cancer. *Breast Cancer Res Treat.* 2012;134(2):701-8.  
19 doi:10.1007/s10549-012-2076-7.
- 20 47. Larsson A, Johansson ME, Wangefjord S, Gaber A, Nodin B, Kucharzewska P et al. Overexpression  
21 of podocalyxin-like protein is an independent factor of poor prognosis in colorectal cancer. *Br J*  
22 *Cancer.* 2011;105(5):666-72. doi:10.1038/bjc.2011.295.
- 23 48. Eberhard J, Gaber A, Wangefjord S, Nodin B, Uhlen M, Ericson Lindquist K et al. A cohort study of  
24 the prognostic and treatment predictive value of SATB2 expression in colorectal cancer. *Br J Cancer.*  
25 2012;106(5):931-8. doi:10.1038/bjc.2012.34.

- 1 49. Maniecki MB, Etzerodt A, Ulhoi BP, Steiniche T, Borre M, Dyrskjot L et al. Tumor-promoting  
2 macrophages induce the expression of the macrophage-specific receptor CD163 in malignant cells.  
3 Int J Cancer. 2012;131(10):2320-31. doi:10.1002/ijc.27506.
- 4 50. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S et al. Reducing waste from  
5 incomplete or unusable reports of biomedical research. Lancet. 2014;383(9913):267-76.  
6 doi:10.1016/S0140-6736(13)62228-X.
- 7 51. Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PM, Korevaar DA et al. Increasing value and  
8 reducing waste in biomedical research: who's listening? Lancet. 2016;387(10027):1573-86.  
9 doi:10.1016/S0140-6736(15)00307-4.
- 10 52. Jankova L, Dent OF, Molloy MP, Chan C, Chapuis PH, Howell VM et al. Reporting in studies of  
11 protein biomarkers of prognosis in colorectal cancer in relation to the REMARK guidelines.  
12 Proteomics Clin Appl. 2015;9(11-12):1078-8. doi:10.1002/prca.201400177.
- 13 53. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ et al. CONSORT 2010  
14 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ.  
15 2010;340:c869. doi:10.1136/bmj.c869.
- 16 54. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D. Does use of the CONSORT Statement  
17 impact the completeness of reporting of randomised controlled trials published in medical journals?  
18 A Cochrane review. Syst Rev. 2012;1:60. doi:10.1186/2046-4053-1-60.
- 19 55. Coppus SF, van der Veen F, Bossuyt PM, Mol BW. Quality of reporting of test accuracy studies in  
20 reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD)  
21 initiative. Fertil Steril. 2006;86(5):1321-9. doi:10.1016/j.fertnstert.2006.03.050.
- 22 56. Smidt N, Rutjes AW, van der Windt DA, Ostelo RW, Bossuyt PM, Reitsma JB et al. The quality of  
23 diagnostic accuracy studies since the STARD statement: has it improved? Neurology. 2006;67(5):792-  
24 7. doi:10.1212/01.wnl.0000238386.41398.30.

- 1 57. Selman TJ, Morris RK, Zamora J, Khan KS. The quality of reporting of primary test accuracy studies  
2 in obstetrics and gynaecology: application of the STARD criteria. *BMC Womens Health*. 2011;11:8.  
3 doi:10.1186/1472-6874-11-8.
- 4 58. Bastuji-Garin S, Sbidian E, Gaudy-Marqueste C, Ferrat E, Roujeau JC, Richard MA et al. Impact of  
5 STROBE statement publication on quality of observational study reporting: interrupted time series  
6 versus before-after analysis. *PLoS One*. 2013;8(8):e64733. doi:10.1371/journal.pone.0064733.
- 7 59. da Costa BR, Cevallos M, Altman DG, Rutjes AW, Egger M. Uses and misuses of the STROBE  
8 statement: bibliographic study. *BMJ Open*. 2011;1(1):e000048. doi:10.1136/bmjopen-2010-000048.
- 9 60. Vandembroucke JP. STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION,  
10 COREQ, QUOROM, REMARK... and CONSORT: for whom does the guideline toll? *J Clin Epidemiol*.  
11 2009;62(6):594-6. doi:10.1016/j.jclinepi.2008.12.003.
- 12 61. Stevens A, Shamseer L, Weinstein E, Yazdi F, Turner L, Thielman J et al. Relation of completeness  
13 of reporting of health research to journals' endorsement of reporting guidelines: systematic review.  
14 *BMJ*. 2014;348:g3804. doi:10.1136/bmj.g3804.
- 15 62. Wang X, Chen Y, Yang N, Deng W, Wang Q, Li N et al. Methodology and reporting quality of  
16 reporting guidelines: systematic review. *BMC Med Res Methodol*. 2015;15:74. doi:10.1186/s12874-  
17 015-0069-z.
- 18 63. Marchionni L, Wilson RF, Wolff AC, Marinopoulos S, Parmigiani G, Bass EB et al. Systematic  
19 review: gene expression profiling assays in early-stage breast cancer. *Ann Intern Med*.  
20 2008;148(5):358-69.
- 21 64. Diamandis EP. The failure of protein cancer biomarkers to reach the clinic: why, and what can be  
22 done to address the problem? *BMC Med*. 2012;10:87. doi:10.1186/1741-7015-10-87.
- 23 65. Kleinert S, Horton R. How should medical science change? *Lancet*. 2014;383(9913):197-8.  
24 doi:10.1016/S0140-6736(13)62678-1.
- 25 66. Altman DG. Making research articles fit for purpose: structured reporting of key methods and  
26 findings. *Trials*. 2015;16:53. doi:10.1186/s13063-015-0575-7.

- 1 67. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A et al. Prognosis research strategy  
2 (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*. 2013;346:e5595.  
3 doi:10.1136/bmj.e5595.
- 4 68. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S et al. Prognosis  
5 Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381.  
6 doi:10.1371/journal.pmed.1001381.
- 7 69. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW et al. Prognosis research  
8 strategy (PROGRESS) 4: stratified medicine research. *BMJ*. 2013;346:e5793. doi:10.1136/bmj.e5793.
- 9 70. PROGNosis REsearch Strategy (PROGRESS) group. <http://progress-partnership.org/>.
- 10 71. STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative.  
11 <http://www.stratos-initiative.org/>.
- 12 72. Sauerbrei W, Abrahamowicz M, Altman DG, le Cessie S, Carpenter J, initiative S. STRengthening  
13 analytical thinking for observational studies: the STRATOS initiative. *Stat Med*. 2014;33(30):5413-32.  
14 doi:10.1002/sim.6265.
- 15 73. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Ioannidis JP. Reproducible Research Practices and  
16 Transparency across the Biomedical Literature. *PLoS Biol*. 2016;14(1):e1002333.  
17 doi:10.1371/journal.pbio.1002333.
- 18 74. Sauerbrei W, Hollander N, Riley RD, Altman DG. Evidence-based assessment and application of  
19 prognostic markers: The long way from single studies to meta-analysis. *Commun Stat-Theor M*.  
20 2006;35(7):1333-42. doi:10.1080/03610920600629666.
- 21 75. Andre F, McShane LM, Michiels S, Ransohoff DF, Altman DG, Reis JS et al. Biomarker studies: a  
22 call for a comprehensive biomarker study registry. *Nat Rev Clin Oncol*. 2011;8(3):171-6.  
23 doi:10.1038/nrclinonc.2011.4.
- 24 76. Altman DG. The time has come to register diagnostic and prognostic research. *Clin Chem*.  
25 2014;60(4):580-2. doi:10.1373/clinchem.2013.220335.

26



# 1 **Supporting Information**

2

3 **S1 Doc: REMARK checklist**

4 **S2 Doc: Eligibility criteria for selection of studies**

5 **S3 Doc: References of selected studies**

6 **S4 Doc: Data extraction form**

7 **S5 Doc: Assessed and discarded items of REMARK checklist – Reasons and definitions**

8

9 **S1 Table: Selected articles over time**

10 **S2 Table: Summary statistics**

11

12 **S1 Data: Analysed data** *(additional excel-file)*

Dear Reviewer,

Thank you very much for reviewing our manuscript. Subsequently, you will find our reply following each of your comments.

Kind regards,  
Peggy Sekula

---

### Reviewers' comments:

Reviewer #1: Well done study, nicely presented; I only have a few comments:

1. is there a reason for the precise time limit that you placed to review the journals you did.

**Reply:** The PRE-study includes publications published mainly in 2006 and few in 2007. The POST-study includes publications published mainly in 2011 and 2012 but in single situations (Journals: *Canc, IJC*) it goes back to 2007. We ensured that there is no overlap in selected manuscripts. The search itself was conducted in March 2013 - therefore the defined period of assessed articles of 2007 to 2012. We did not repeat the search as we already had enough publications but also due to limited manpower. No change in the manuscript.

2. I understand that you chose the journals that represent those in the first survey, but would there be value in surveying other journals with high impact factor in the field (JNCI, JAMA Oncol, Lancet Oncol) in addition to those already chosen? For example, the latter two didn't exist when the first survey was done. however, this may be a new study for a subsequent article.

**Reply:** The project focused on the assessment of the potential improvement in reporting quality. For this reason, we attempted to select articles from the same journals in order to sensibly compare results with the past. Since we did not identify enough articles in the preferred journals we already included 2 other journals knowing that this limits our conclusions regarding comparability of results. Of course, we could have included other journals as well. In this situation, however, the focus would be rather on quality assessment in general and between journals. Currently we can only hypothesize that reporting quality will be similarly poor in other related journals. Additionally, due to limited manpower in this project, such assessment could be only addressed in a new study. Parts of these explanations were already mentioned in the first version (see p.6 Material and Methods 2nd paragraph und p. 16 Discussion: 1 Limitations of study). No change in the manuscript.

3. It would be a great service if the REMARK authors would prepare an automated, electronic checklist for authors that includes each of the points and provides appropriate questions as to whether that point is covered or not. Perhaps such a project could be brought up in the Discussion?

**Reply:** Thank you for your suggestion of an 'automated, electronic checklist'. It is an excellent idea as it may provide extra help in the future, in addition to the free downloads available for current REMARK checklist and extensive guidance already published. We are therefore more than willing to include your suggestion in the conclusions of our manuscript (p. 19). However, we consider it more as the task and the privilege of the authors of the REMARK checklist (two of us (DGA and WS) belong to them). It will be discussed in this group and it is likely that an electronic checklist on the web providing explanations to each item in pop up windows will be available soon. However, it is unclear if journals are willing to add such an instrument to their submission guidelines. Anyhow, it is worth a try.