

Doug Altman: Driving critical appraisal and improvements in the quality of methodological and medical research

Willi Sauerbrei¹ | Martin Bland² | Stephen J. W. Evans³ | Richard D. Riley⁴ | Patrick Royston⁵ | Martin Schumacher¹ | Gary S. Collins⁶

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

² Department of Health Sciences, University of York, York, UK

³ Medical Statistics Department, London School of Hygiene and Tropical Medicine, London, UK

⁴ Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Staffordshire, UK

⁵ MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, University College London, London, UK

⁶ Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford, UK

Correspondence

Willi Sauerbrei, Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, 79104 Freiburg, Germany.
Email: wfs@imbi.uni-freiburg.de

Abstract

Doug Altman was a visionary leader and one of the most influential medical statisticians of the last 40 years. Based on a presentation in the “Invited session in memory of Doug Altman” at the 40th Annual Conference of the International Society for Clinical Biostatistics (ISCB) in Leuven, Belgium and our long-standing collaborations with Doug, we discuss his contributions to regression modeling, reporting, prognosis research, as well as some more general issues while acknowledging that we cannot cover the whole spectrum of Doug’s considerable methodological output. His statement “*To maximize the benefit to society, you need to not just do research but do it well*” should be a driver for all researchers. To improve current and future research, we aim to summarize Doug’s messages for these three topics.

KEYWORDS

prognosis research, regression modeling, reporting

1 | INTRODUCTION

Doug Altman, who passed away on June 3, 2018, was one of the most influential medical statisticians of the past 40 years. His first reaction to this statement would probably have been, “*But how can you make such a strong statement? Where is the reliable and accessible evidence?*” Although collecting convincing empirical evidence for strong statements is often difficult and time-consuming, in this case the argument is easily made.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Biometrical Journal* published by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

Doug's 1991 textbook *Practical Statistics for Medical Research* has sold more than 50,000 copies until 2015 (<https://www.ndorms.ox.ac.uk/news/professor-doug-altman-receives-bmj-lifetime-achievement-award>), making it one of the most popular in this field (Altman, 1991a). In 1994, he published an editorial in the *British Medical Journal* (BMJ) entitled "The scandal of poor medical research – we need less research, better research and research done for the right reasons" (Altman, 1994). This short article has received relatively few citations compared with his other articles, at only several hundred, but has profoundly affected the medical research community. There were earlier articles about the poor quality of methodology and reporting in medical research, and Doug's editorial appeared at the same time as many other critical papers. But it was Doug's clear message that resonated with the community: "*As the system encourages poor research it is the system that should be changed. We need less research, better research, and research done for the right reasons. Abandoning using the number of publications as a measure of ability would be a start.*"

In the years that followed, Doug stressed the key role of reporting guidelines and was one of the authors of the *Lancet* series "Research: increasing value, reducing waste" published in 2014 (<http://www.thelancet.com/series/research>). In an introductory comment, Kleinert and Horton (2014) stated "... a far broader question should be posed: how should the entire scientific enterprise change to produce reliable and accessible evidence that addresses the challenges faced by society and the individuals who make up those societies?" There is no longer debate about whether medical science needs to change. The question now is *how* it should change. More than 25 years ago, Doug strongly expressed the need to change the system (Altman, 1994). In the years that followed, he pressed for improvements across many areas, including primary studies and systematic reviews. Unfortunately, flaws in the application and interpretation of statistical analyses are still common (Page et al., 2018).

Doug disagreed with the idea that the number of publications one amasses indicates the quality of one's research. However, the number of citations garnered by his papers does show the impressive reach of his work: as of January 16, 2020, Google Scholar records around 470,000 citations for his papers. About 70 of these papers have been cited more than 1,000 times. Even more astoundingly, Doug's papers are currently the most cited in the *Lancet*, *BMJ*, *Annals of Internal Medicine*, *Statistical Methods in Medical Research*, *PLoS Medicine*, *Ultrasound in Obstetrics Gynecology*, and *Journal of Biopharmaceutical Statistics*.

Reliable evidence for our statement about Doug's outstanding scientific contribution comes from the *BMJ*, for which Doug served as Chief Statistical Editor for more than 20 years. In 2015, the *BMJ* honored Doug with their Lifetime Achievement Award, saying: "*Doug Altman has been awarded the BMJ Lifetime Achievement Award in recognition of his outstanding contribution to the improvement of the scientific and medical research literature. Professor Altman is one of the world's leading experts in health research methodology, statistics and reporting and has spent his career working to improve transparency in the conduct and reporting of health research. Over the years Professor Altman has led or been involved in developing many of the reporting guidelines listed on the EQUATOR website... Altman has done more than anybody to raise the standards of medical publication and in the process has transformed the role of statistician from number cruncher to custodian of important but often neglected values.*" (<http://www.equator-network.org/2015/05/07/professor-doug-altman-awarded-bmj-lifetime-achievement-award/>)

Doug was a brilliant collaborative researcher who worked with many national and international colleagues on an enormous range of methodological and medical projects. Several published tributes give more detail than we could possibly manage here (Collins, 2019; Deeks et al., 2018; Evans, 2018; Matthews, Chalmers, & Rothwell, 2018; Rennie, 2019; Trivella, 2019; Watts, 2018). Although he was a pioneer of many novel statistical methods, Doug emphasized that a statistician's biggest impact was often in ensuring adherence to basic statistical good practice. He championed practices, such as focusing on effect estimates and confidence intervals rather than *p*-values, and analyzing continuous variables on their continuous scale rather than categorizing them.

Doug's collaborations were not just broad in topic. As noted at the 2018 Royal Statistical Society Annual Conference tribute, (<https://www.youtube.com/watch?v=f6k6LlqbqGA>), Doug published with over 1,600 coauthors. The many coauthors came from many countries with the top 25 shown in Figure 1 (bottom), so even for the 25th country, South Africa, eight of his papers were with authors from there.

In July 2019, one of us (Willi Sauerbrei) gave a presentation in the "Invited session in memory of Doug Altman" at the 40th Annual Conference of the International Society for Clinical Biostatistics (ISCB) in Leuven, Belgium. For more than 25 years, Willi collaborated with Doug and many other colleagues. After discussing the intended talk with some of these collaborators, Willi decided to concentrate on regression modeling, reporting, and prognosis research, as his joint work with Doug in these areas has had a major influence on his professional life. To broaden the view and illustrate Doug's profound influence on the current state of research, this article also includes contributions from other colleagues who worked with Doug in at least one of these areas. For each of the topics discussed, we attempt to summarize the

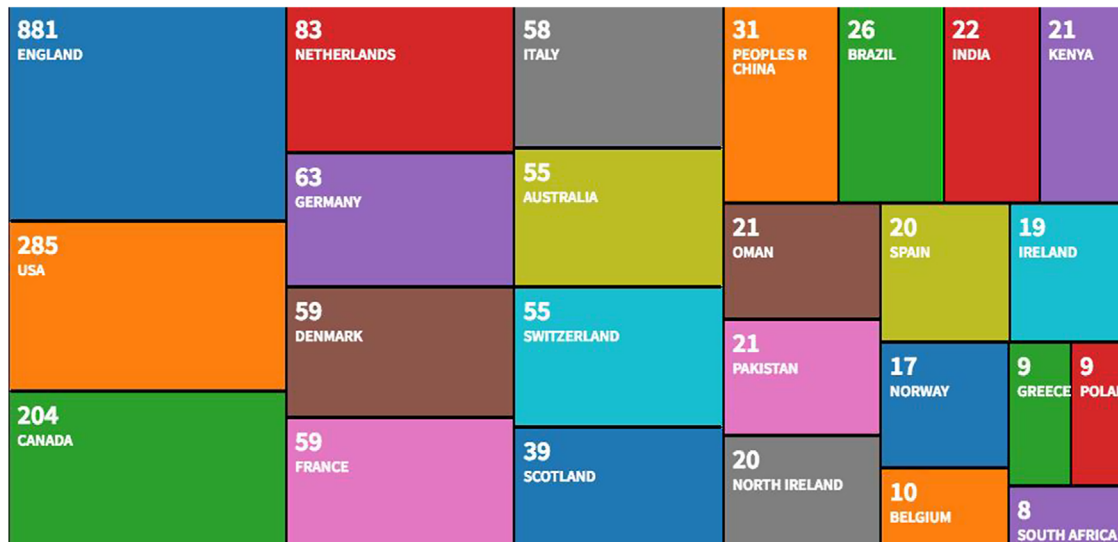


FIGURE 1 (Top) Doug's most frequent coauthors, with size proportional to frequency, and (bottom) the countries that his coauthors came from, with area proportional to frequency

messages and advice that Doug and his collaborators have given to the research community. Each topic is very broad, and we consider only specific issues in which we have jointly published with Doug. For further topics and details, we refer the reader to textbooks written or edited by Doug (Altman, 1991a; Altman, Machin, Bryant, & Gardner, 2000; Gore & Altman, 1982; Moher, Altman, Schulz, Simera, & Wager, 2014; Riley, van der Windt, Croft, & Moons, 2019) and to references given in tributes mentioned above.

2 | REGRESSION MODELING

Doug's joint work with Willi and Martin Schumacher on regression modeling started when Doug sent Martin a manuscript on bootstrap investigations for comments at the beginning of 1988. In the published version, Altman and Andersen (1989) thanked Martin, Willi, Patrick Royston, and others for helpful comments on an earlier version. We could not have foreseen

that Doug and Patrick's later work (Royston & Altman, 1994) would lead to Patrick and Willi's joint work on multivariable fractional polynomials (Royston & Sauerbrei, 2008).

Continuous variables play a key role in all areas of science in which empirical data are analyzed, and categorization has been commonly used to handle continuous variables in multivariable regression models. In the 1990s, Doug and others pointed to the critical issues raised by using cutpoints to categorize continuous variables (Altman, 1991b; Altman, Lausen, Sauerbrei, & Schumacher, 1994). To improve this unfortunate situation, Doug and Patrick proposed the fractional polynomial approach to determine the functional form of a continuous variable (Royston & Altman, 1994).

2.1 | Bootstrap investigations of model stability

A major part of Doug's methodological work was concentrated on regression modeling. All of his contributions were motivated by concrete problems arising when analyzing a particular data set, both from projects where he was the statistician and from papers that he discovered in published medical research. Doug centered his teaching and writing about statistical methods in real-world examples. His 1991 book illustrates this approach, which may be one of the reasons for its widespread use in the medical research community (Altman, 1991a; see also Section 6).

One example of these real-world problems was in a multinational placebo-controlled double-blind randomized trial in patients with primary biliary cirrhosis (Christensen et al., 1985). The trial showed a significant treatment benefit if (and only if) partially unbalanced covariates were adjusted for. The question arose whether this result was sensitive to the choice of regression model for the adjustment of the treatment effect and whether the data-dependent choice of covariates would be stable when replicated. When a second study is not available, bootstrap resampling can be used to investigate model stability. Building on work by Chen and George (1985), Doug and Per Kragh Andersen used bootstrap resampling by integrating various variable selection approaches, so that an adjustment model was chosen in every bootstrap sample (Altman & Andersen, 1989). They were able to provide clear evidence that the original model chosen for the trial analysis and the resulting treatment effect estimates were valid. Willi and Martin extended this procedure by using the summary results of bootstrap resampling to create a stable regression model, which also addressed the correlation between covariates (Sauerbrei & Schumacher, 1992).

At the time, these bootstrap resampling approaches were still rare. The importance of stability investigations is now more widely recognized (Heinze, Wallisch, & Dunkler, 2018), and resampling procedures are on the way to being routinely used to investigate the stability of regression models, particularly when high-dimensional covariate data are involved.

2.2 | Cutpoints to categorize continuous variables

Doug was always well acquainted with the medical literature, noticing trends in the application of specific statistical methodology—and serious shortcomings—earlier than others. His work on categorizing continuous variables is one example.

In 1991, Doug wrote a letter to the editor of the *British Journal of Cancer* (Altman, 1991b) to draw attention to issues in an article on the role of the fraction of cells in the S-phase (SPF) for the prognosis of patients with breast cancer (Sigurdsson et al., 1990). Doug observed that the authors had tried every possible cutpoint when categorizing SPF and reported the “optimal” cutpoint leading to the most significant effect of SPF on disease-free survival, that is, the minimum p -value.

At the same time, Martin Schumacher and Berthold Lausen derived the asymptotic distribution of maximally selected rank statistics (Lausen & Schumacher, 1992), building on the work by Miller and Siegmund (1982). As the maximally selected log rank test is associated with the minimum p -value, this corresponds to the situation that Doug observed in the aforementioned article. The type 1 error rate was, therefore, about 50%, not the 5% implied by the reported significance test. By naively applying a cutoff that gives a minimum p -value, researchers can inflate a factor's prognostic effect, when in reality it may not be prognostic at all. The minimum p -value can be corrected relatively simply by taking the maximization of the test statistic or the minimization of the p -value into account.

During a meeting, Doug, Berthold Lausen, Martin, and Willi decided that this problem should be brought to the attention of a broader audience, particularly clinicians. We joined forces and wrote a manuscript exemplifying the problem using data from a study of prognostic factors for primary breast cancer, using SPF as the factor of particular interest. We searched for studies that dichotomized the continuous SPF value into two groups and investigated whether recurrence-free survival times differed between the groups. We found 19 different cutpoints used for dichotomization, ranging from

2.6 to 15.0. We also used this study to show that it is (nearly) always possible to define a “suitable” study population and cutpoint that indicates that a factor has a significant prognostic effect. The resulting commentary in the *Journal of the National Cancer Institute* (Altman et al., 1994) has often been used by editors to reject papers on prognostic factors that use the minimum p -value.

However, despite this important early work highlighting the problems of categorizing continuous variables, this poor practice continued to appear in the published literature. Doug, Willi, and Patrick wrote a paper in response that again summarized the disadvantages of categorization and presented other approaches that make dichotomization of continuous predictors unnecessary (Royston, Altman, & Sauerbrei, 2006). With the medical community in mind, Doug also published a statistics note in the *BMJ* with Patrick, entitled “The cost of dichotomizing continuous variables” (Altman & Royston, 2006). These papers remain highly pertinent, as sadly dichotomization is still rife in medical research.

2.3 | Fractional polynomials

Royston and Altman (1994) introduced fractional polynomials as a useful extension of polynomial regression that could be used to model the relationship between a continuous variable and an outcome. They proposed an extended family of curves whose power terms are restricted to a small predefined set of integer and noninteger values. The powers are selected so that conventional polynomials are a subset of the family. A suitable function selection procedure can be used to check whether a linear function (in general the default) is adequate or whether a nonlinear fractional polynomial function substantially improves model fit (Royston & Sauerbrei, 2008).

Although read to the Royal Statistical Society, the original paper was largely ignored by the statistical and medical communities. Spline-based approaches were preferred as they are much more flexible for deriving a functional form for one continuous variable. Doug’s interests switched to other methodological issues, such as systematic reviews, meta-analysis, and reporting. He did not further promote fractional polynomial methodology beyond one more paper with Patrick (Royston & Altman, 1997) on using fractional polynomials to approximate smooth, continuous mathematical functions.

However, Willi realized that fractional polynomial methodology could be an important component in a multivariable context. For many years, he had worked on variable selection, considering how continuous variables are handled to be a key problem. Patrick and Willi combined their work and derived the multivariable fractional polynomial algorithm and several extensions (Royston & Sauerbrei, 2008; Sauerbrei & Royston, 1999). Although other researchers have considered extending the original proposal of the fractional polynomial power terms, practical experience has shown that the class of fractional polynomial functions proposed by Royston and Altman (1994) is an excellent choice for the statistical analysis of many data sets.

2.4 | Age-related reference intervals

Reference intervals are predefined centiles of a clinically relevant variable Y . Loosely speaking, an individual is considered normal if Y lies within the reference interval and potentially abnormal otherwise. Statistically, reference intervals are relatively simple to construct when Y is considered alone but more challenging when the distribution of Y is associated with a continuous predictor, often age.

Traditionally, age-related reference intervals are estimated by imposing cutpoints on age and applying univariate methods of centile estimation to the resulting age groups. The approach has several obvious and severe flaws.

The preferable alternative is to keep age continuous. Simple regression methods will fail to estimate reference intervals accurately when the mean and higher moments of Y depend linearly or nonlinearly on age. Doug described a method based on absolute residuals to estimate the age-related mean and SD of Y when the conditional distribution of Y given age is normal (Gaussian) (Altman, 1993). In collaboration with Dr Lyn Chitty (University College London), Doug used the absolute-residuals/FP method just mentioned to develop “charts” (graphs depicting gestational age-related reference intervals) for a wide variety of fetal measurements (Altman & Chitty, 1994).

The general idea of extending conventional polynomials to fractional polynomial models arose from Doug’s approach to estimating the gestational age-related mean of fetal size. For instance, he found that restricted cubic functions of the form $\beta_1 x + \beta_2 x^3$ provided an excellent, smooth fit to several measurements of fetal size. Discussion with Patrick Royston on this topic led to joint work to develop and formalize fractional polynomial models and their inference (Royston & Altman, 1994).

During the last decade of his life, Doug made substantial contributions to the design and analysis of the International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st, a WHO-funded study of age-related reference intervals for fetal size and growth, see <https://www.wrh.ox.ac.uk/research/intergrowth-21st> and references therein).

One of Doug's key principles of statistical modeling is clearly expressed in the statistical analysis section of the most cited paper from this Consortium (Villar et al., 2014), 740 citations in Google Scholar (April 29, 2020). Authors discuss several statistical methods, which may be used to construct standards to assess the size and weight of a newborn. They state as criteria "Our aim was to produce centiles that change smoothly with gestational age and that maximise simplicity without compromising model fit." Such charts are routinely used for monitoring fetal growth and development worldwide.

In two recent papers, standards for statistical methodology (Ohuma & Altman, 2019) and fetal growth (Papageorgiou et al., 2018) were proposed. Doug's name appears as a coauthor of many research reports on the INTERGROWTH-21st project, four of which have more than 100 citations in Google Scholar (April 29, 2020).

2.5 | COSMIC

During the 1990s, Doug, Patrick, Willi, Martin Schumacher, and Hans van Houwelingen focused on deeper issues in regression modeling. Doug referred to these discussions as COSMIC (Cooperation On Statistical Modelling In Cancer)—he was quite extraordinary in coming up with suitable acronyms to describe projects and often devoted considerable energy to the process. Several papers resulted from this collaboration. Presentations under the COSMIC name were given, but no paper was coauthored by all five members. However, the COSMIC discussions profoundly influenced all of their work.

Collaborative projects included building reliable and clinically valuable prognostic models in common types of cancer; validating prognostic classification schemes in node-negative breast cancer patients; defining basic requirements for a meta-analysis to investigate the prognostic effect of the different measurements of the plasminogen activation system in breast cancer; and a simulation study to compare approaches for model building in survival data with time-varying effects.

Only one paper clearly mentions the COSMIC group. In their highly cited article "What do we mean by validating a prognostic model?" Altman and Royston (2000) acknowledged "*We thank Hans van Houwelingen, Willi Sauerbrei and Martin Schumacher (the other members of the COSMIC group), Marc Buyse and Jeremy Wyatt for helpful comments and discussions.*" The group members agreed that considering both statistical and clinical validity was sensible, but differed on a few specific issues. Consequently, no paper was written under the name COSMIC.

Many papers authored or coauthored by a COSMIC member were influenced by discussions between the five members and the PhD students who joined their meetings. The main topics of a 2-day meeting in 2005 illustrate the group's broad interests: pitfalls in interim analysis on longitudinal and survival data, reduced rank hazard regression versus frailty and cure models, methods for estimating length of hospital stay (time-dependent bias), prognostic models in breast cancer, the multivariable fractional polynomial-time algorithm, investigations of neural nets, model selection uncertainty, cross validation and model selection, and a plea for good statistical practice when using high-dimensional data for classification and prediction.

We talked about drafting a COSMIC position paper on some of these issues, but never managed it. However, COSMIC discussions suggested that work was needed to compare statistical procedures and formulate guidance for design and analysis. The discussions helped motivate the project that became the STrengthening Analytical Thinking for Observational Studies (STRATOS) initiative (see Section 5.4), which included all five COSMIC members.

2.6 | Mistakes in the interpretation of analyses

An error of interpretation led to Doug's single most highly cited paper (Bland & Altman, 1986) with 45,483 citations on Google Scholar and 33,659 on Web of Science (April 21, 2020). Martin Bland told Doug about a paper which had used correlation as a measure of the agreement between two methods of measuring cardiac stroke index. Doug replied that he had come across the same thing in blood pressure measurement. They agreed that a correlation coefficient could not measure agreement: if one measurement were exactly twice the size of the other, for example, the correlation coefficient would be 1.0, but the measurements would not agree. Doug looked for other examples and found two other inappropriate methods of analysis, by paired t test and regression, testing the slope against 1.0, ignoring the regression dilution effect. Martin drafted an introduction and discussion, including a mathematical appendix. However, what should be done? They

decided to start with the differences between measurements by the two methods. Having a set of differences, any statistician would calculate the mean and standard deviation and it was a short step to saying that two (or the more accurate figure, 1.96, preferred by Doug) standard deviations above and below the mean would provide a range within which 95% of differences would lie, giving limits for the second measurement. (For the method to be valid, certain assumptions about the distribution of the differences were of course needed.) Standard errors and confidence intervals for the limits of agreement completed the paper. As a check on the constancy of mean and standard deviation, they suggested plotting difference against the mean of the two measurements by the two methods, which they thought was a standard approach for paired *t* tests. They then proposed adding the 95% limits of agreement to the plot.

The idea was submitted for a conference of the Institute of Statisticians, now part of the Royal Statistical Society. Neither Doug nor Martin had ever spoken at a statistical meeting before. Surprisingly, this simple approach was not known. Doug and Martin wrote a paper, which was published in *The Statistician*, the journal of the Institute of Statisticians (Altman & Bland, 1983). It was well received by statisticians, but ignored by researchers, who carried on correlating. Colleagues suggested writing a version with a worked example for medical researchers. Martin collected some data by touring his department with a peak expiratory flow meter and a mini meter, making two measurements by each method in random order. He added himself and Doug to the data and also his parents and parents-in-law. The data were analyzed and formed the core of a paper published by the *Lancet* (Bland & Altman, 1986).

To their great surprise and pleasure, the *Lancet* paper began to be cited, rapidly becoming a Citation Classic (Bland & Altman, 1992), and eventually the most highly cited paper in the *Lancet*. In 2014, it was reported to be the 29th most highly cited paper in any journal in any field (Van Noorden, Maher, & Nuzzo, 2014). Although the *Statistician* paper also began to be cited, the *Lancet* paper was seen as the one to cite and the approach became known as the Bland–Altman method. In response to questions they were sent, Martin and Doug subsequently published further papers on the topic. A paper discussing several extensions of the basic approach has more than 7,000 citations in Google Scholar (Bland & Altman, 1999). More information on the history of this work is available (Bland & Altman, 1995a).

3 | REPORTING

Doug became interested in the quality of statistical reporting early in his career. While he was in his first post, at St. Thomas's Hospital Medical School, London, he and Martin Bland submitted a letter to the *Lancet* complaining about a paper which had described 321 people as having been allocated to groups “more or less randomly” (Clarke & Campbell, 1975). They argued that allocation is either random or it is not; there is no intermediate state. The letter was not published, but the point they were trying to make (albeit not very well) was a general one about research methods. They did get a letter published criticizing statistical methods used in a *Lancet* article (Dritz et al., 1977), where association had been mistaken for causation (Bland & Altman, 1977). The letter was the first of more than 100 joint scientific publications (see Sections 2.6 and 5.1).

In 1993, a group of medical statisticians and others involved in trials and their reporting met in Ottawa, Canada, led by David Moher and Stuart Pocock. Stuart had published an article (Pocock, Hughes, & Lee, 1987) on some of the problems in how trials were reported, after he, Doug, Sheila Bird (nee Gore), and the late Martin Gardner had suggested general guidelines in an earlier paper (Altman, Gore, Gardner, & Pocock, 1983). The Ottawa meeting resulted in the SORT guidelines for reporting clinical trials (Andrew et al., 1994). The initiative merged with another to form the Consolidated Standards of Reporting Trials (CONSORT) group, which published the first CONSORT reporting guideline in 1996. Although Doug was not an author, he was a cosupervisor of coauthor Ken Schultz's PhD (1994) and published with Ken and others on reporting standards in trials and observational studies.

3.1 | CONSORT and the EQUATOR Network

The CONSORT group met again in 1999. After careful revision, it published an updated version of the guidelines in 2001 (Moher, Schulz, & Altman, 2001). The paper states “David Moher, Ken Schulz, and Doug Altman participated in regular conference calls, identified participants, contributed in the CONSORT meetings, and drafted the manuscript.” The CONSORT group has been extraordinarily productive, not only in their own activities but also in encouraging wider involvement with many others in different countries. The result has been further revisions of CONSORT, many extensions for specific trial designs and clinical areas, and translations.

The success of CONSORT led to reporting guidelines for other study designs, including the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for observational research (von Elm et al., 2007), the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for systematic reviews (Moher, Liberati, Tetzlaff, & Altman, 2009), and extensions. Doug was involved in the development of many of these guidelines, including STROBE and PRISMA. As has been done for many of the high-profile reporting guidelines, the PRISMA statement was simultaneously published in six journals. It is currently Doug's most highly cited paper, with well over 60,000 citations according to Google Scholar. Many of the CONSORT extensions are also highly cited, with extensions for cluster trials (Campbell, Elbourne, & Altman, 2004), pragmatic trials (Zwarenstein et al., 2008), noninferiority trials (Piaggio, Elbourne, Altman, Pocock, & Evans, 2006), and reporting harms (Ioannidis et al., 2004) all garnering more than 1,000 citations.

Despite CONSORT's success, uptake of reporting guidelines by the wider research community was initially slow. Largely through Doug's initiative, the Enhancing the QUALity and Transparency Of health Research (EQUATOR—another great acronym!) Network was born in 2006 to promote reporting guidelines (<https://www.equator-network.org/>). The official launch meeting photograph shows Doug wearing a tie, a most unusual occurrence! (<https://www.equator-network.org/about-us/history/>). The EQUATOR Network holds a comprehensive database of reporting guidelines for health research for any study design, numbering over 400 in early 2020. A summary of its goals and activities is given in Simera et al. (2010). Of course, many researchers develop reporting guidelines. To improve and partly standardize the process, it is important that guidance for such developers exists (Moher, Schulz, Simera, & Altman, 2010). As already mentioned in the introduction, Doug was keen on the availability of reliable empirical evidence on an issue of interest and motivated researchers to work with him on systematic reviews. For randomized controlled trials with statistically nonsignificant primary outcomes, Boutron, Dutton, Ravaud, and Altman (2010) showed that the reporting and interpretation of findings were frequently inconsistent with the results.

3.2 | REMARK

Simon and Altman (1994) published a seminal paper about statistical aspects of prognostic factor studies, followed by a paper about methodological challenges (Altman & Lyman, 1998). These papers influenced general thinking about the poor quality of prognostic research. Despite years of research and hundreds of published reports on tumor markers in oncology, only a small number of markers have proven clinically useful (Hall et al 1999; Hayes et al., 1996; McShane et al., 2000). To rectify this unfortunate situation, in 2000, the US National Cancer Institute (NCI) and the European Organization for Research and Treatment of Cancer (EORTC) organized the First International Meeting on Cancer Diagnostics (titled From Discovery to Clinical Practice: Diagnostic Innovation, Implementation, and Evaluation). The purpose of the meeting was to discuss issues, accomplishments, and barriers in the field of cancer diagnostics.

Inadequate reporting was identified as a key issue. Doug (for EORTC) and Lisa McShane (for NCI) were selected as chairs of a project to derive guidelines for reporting tumor marker prognostic studies. Doug invited Willi to join a small group of researchers to work on these guidelines. Despite being skeptical about work on guidance for reporting, Willi joined the project. The REporting recommendations for tumor MARKer prognostic studies (REMARK) guidelines were initially published in 2005 in five journals (*Journal of the National Cancer Institute*, the *British Journal of Cancer*, the *European Journal of Cancer*, the *Journal of Clinical Oncology*, and *Nature Clinical Practice Oncology*), then again in 2006 in a further two journals (*Breast Cancer Research and Treatment* and *Experimental Oncology*) (McShane et al., 2005).

To enlarge on the 20-item REMARK checklist, four of the coauthors created an Explanation and Elaboration (E&E) paper. Doug had had experience with E&E papers for other reporting guidelines. Such papers explain each item of the reporting guideline in detail and give published examples of good reporting. Putting together, the REMARK E&E was no simple task. The aim was to provide a comprehensive overview to educate researchers on good reporting and provide a valuable compendium of issues to consider when designing, conducting, and analyzing tumor marker studies and prognostic studies in medicine in general. For example, Item 10 of the checklist states that authors should report all statistical methods used. The E&E lists several aspects of analysis that could be reported under eight subheadings. For some methodological issues, additional details were provided in five boxes, for example, missing data, subgroups, and interactions. The authors also developed the two-part REMARK profile for structured reporting of all analyses planned and undertaken (see Section 5.3 for further details). Before submitting the paper for publication, Doug insisted on sending it to several experienced colleagues for comment. The REMARK E&E project took several years to complete and was finally published in 2012 (Altman, McShane, Sauerbrei, & Taube, 2012).

Not only did Doug develop reporting guidelines, but he was also involved in many systematic reviews evaluating the reporting quality of primary research against reporting guidelines. He was a driving force behind a project illustrating reporting quality in tumor marker studies before REMARK was published (Mallett, Timmer, Sauerbrei, & Altman, 2010). The study was repeated some years later to assess the guideline's influence on the quality of reporting (Sekula, Mallett, Altman, & Sauerbrei, 2017). Alongside this comparison over time, the authors also compared the reporting quality of papers that did and did not cite REMARK, matched by journal and year. Ten of the 20 items were used to define a summary score of reporting quality. Although there was no worthwhile difference between articles that did and did not cite REMARK, there was a slight improvement in reporting quality over time. Serious weaknesses in the quality of reporting remained. Not all the papers that cited REMARK actually applied it correctly.

3.3 | TRIPOD

Doug continued work on improving the reporting of prognostic research beyond tumor marker studies and REMARK. He was involved in numerous systematic reviews evaluating the quality of reporting of prognostic models (Burton et al., 2004; Mallett, Royston, Dutton, Waters, & Altman, 2010; Mallett, Royston, Waters, Dutton, & Altman, 2010). In response to poor reporting described in these reviews and others by Gary Collins and Karel Moons, the TRIPOD initiative was established in 2010, led by Gary, Karel, Johannes Reitsma, and Doug. Once again, Doug coined the acronym, TRIPOD, for the initiative, standing for Transparent Reporting of a multivariable prediction models for Individual Prognosis Or Diagnosis.

Between 2010 and Doug's last overseas visit in April 2018, the four of us met two to three times each year, alternating between Oxford and Utrecht. The TRIPOD initiative aimed to provide guidance on reporting prediction model studies. In 2015, we published the TRIPOD statement simultaneously in 11 journals (*Annals of Internal Medicine*, *British Journal of Obstetrics and Gynecology*, *BMC Medicine*, *BMJ*, *British Journal of Cancer*, *British Journal of Surgery*, *Circulation*, *Diabetic Medicine*, *European Journal of Clinical Investigation*, *European Urology*, and *Journal of Clinical Epidemiology*; Collins, Reitsma, Altman, & Moons, 2015), with editorials and commentaries appearing elsewhere.

Accompanying the checklist was a 73-page E&E paper with 532 references that provided the rationale for each item in the checklist, evidence of poor reporting, and examples of good reporting (Moons et al., 2015). Although the aim of the TRIPOD statement was to improve reporting, the E&E document also aimed to improve methodological quality. It contained methodological considerations, highlighting statistical practices to avoid and consider (e.g., handling missing data, appropriate treatment of continuous variables, and preferred internal validation procedures). Subsequent activities by the TRIPOD group initiated with Doug's involvement include strategies and guidance to assess and monitor adherence to TRIPOD (Heus et al., 2018; Heus et al., 2019), TRIPOD for abstracts, TRIPOD for electronic health records data and individual participant meta-analysis, and, more recently, TRIPOD for machine learning (Collins & Moons, 2019).

The success of the TRIPOD initiative, now also including Richard Riley, led to numerous methodological collaborations on sample size, how to handle continuous predictors, how to handle treatments, external validation, and systematic reviews of prediction model studies. Doug played a prominent role in generating ideas and providing the critical feedback and insights that only someone of his stature and experience could give. His legacy within the group will continue to be felt and appreciated into the future.

4 | PROGNOSIS RESEARCH

Doug's longstanding interest in improving prognostic (or prediction) model research extended beyond fractional polynomials (Section 2.3) and reporting (Sections 3.2 and 3.3) to include sample size (Collins, Ogundimu, Cook, Manach, & Altman, 2016; Ogundimu, Altman, & Collins, 2016; van Smeden et al., 2016; van Smeden et al., 2019), handling of missing data (Burton & Altman, 2004), and validation (Royston & Altman, 2013). In the 1990s, he coauthored articles signaling problems in the design, analysis, and reporting of prognosis research, often focusing on oncology (Altman & Lyman, 1998; Simon & Altman, 1994; Wyatt & Altman, 1995). Sadly, over 20 years later, many of the issues identified are still apparent in prognosis research, particularly small sample sizes, inappropriate analysis methods, and poor and selective reporting. Doug would often share examples of poor practice with his colleagues, using them as fuel to keep pushing forward messages to improve the field.

One of Doug's biggest concerns was the effect of prognosis studies on subsequent systematic reviews, as he feared that poor quality and biased reporting of primary studies would limit the conclusions of subsequent systematic reviews aiming

to summarize multiple prognosis studies. Subsequent empirical evaluations have shown his concerns were justified, with many systematic reviews recommending improvements to primary research studies (Riley et al., 2013), rather than recommending using existing prognostic findings in clinical practice. Doug's much earlier statement (Altman, 2001) is still highly pertinent: "*As a consequence of the poor quality of research, prognostic markers may remain under investigation for many years after initial studies without any resolution of the uncertainty. Multiple separate and uncoordinated studies may actually delay the process of defining the role of prognostic markers.*"

Starting in 2004, Doug and colleagues, including Richard, initiated the Cochrane Prognosis Methods Group (Riley et al., 2007). The group brought together researchers and clinicians with an interest in generating best evidence to improve the pathways of prognostic research and facilitate evidence-based prognosis results to inform research, service development, policy, and practice (<https://methods.cochrane.org/prognosis/>). Doug worked with the group to deliver guidance for conducting reviews of overall prognosis, prognostic factors, and prognostic models (Debray et al., 2017; Riley, Moons et al., 2019).

In recent years, Doug was delighted that exemplar prognosis reviews were being initiated by Cochrane, some of which have now been published in the Cochrane library. These reviews are a testament to Doug's drive and foresight. He was also the senior researcher in one of the first individual participant data (IPD) meta-analyses to examine the prognostic value of a biomarker. Using IPD from published and unpublished studies, the work showed no clear evidence that micro-vessel density was a prognostic marker in patients with lung cancer, which contradicted a previous systematic review and meta-analysis based only on published results (Trivella, Pezzella, Pastorino, Harris, & Altman, 2007). Doug also used the project to highlight the challenges of obtaining IPD (Altman, Trivella, Pezzella, Harris, & Pastorino, 2006) and encouraged greater data-sharing initiatives.

Another major contribution to the prognosis field was the PROGRESS (PROGnosis RESearch Strategy) initiative, led by Doug, Richard, and Harry Hemingway. This group published a series of papers in the *BMJ* and *PLoS Medicine* that outlined a framework for prognosis research under four themes: overall prognosis, prognostic factors, prognostic models, and predictors of treatment effect (Hemingway et al., 2013; Hingorani et al., 2013; Riley et al., 2013; Steyerberg et al., 2013). The aim was to help researchers appreciate how each form of prognosis research can be used to inform and improve patient outcomes, and to identify good practice for the design, analysis, and interpretation of such studies. The PROGRESS group's work laid the foundation for the book *Prognosis Research in Healthcare: concepts, methods and impact*, published by Oxford University Press in 2019 (Riley, van der Windt et al., 2019). Doug coauthored three chapters, and the book was dedicated to him ("*For Doug, our inspiration to improve prognosis research*"). However, much work remains. We encourage researchers to take up Doug's baton and strive for the high-quality prognosis studies he so desired to see.

5 | DOUG'S MESSAGES

For the three topics discussed, we illustrate the influence and the messages of Doug on different time scales. This graph shows impressively that it takes a long time to make changes in the research culture. There were 20 years between Doug's "scandal" letter and the *Lancet* series "Research: increase value, reduce waste." About 30 years ago, he wrote about problems caused by categorizing continuous variables. Meanwhile, many papers have been published on this issue but categorization is still a very popular approach in many analyses. Nowadays, the EQUATOR network is well accepted but Doug and others struggled many years to get funding for a center working on reporting guidelines. A key step was the establishment of the Centre for Statistics in Medicine (CSM) in Oxford, in September 1995.

The time scale is dominated by joint work Doug did with one or more of us (Figure 2). Therefore, it also illustrates a specific type of publication bias. However, the main aim is to bring Doug's messages for these topics to a broader audience. Specifically, his more recent work may not be known to many researchers. The four sections discuss aspects of medical research that Doug particularly focused on and contributed to in a major way because he perceived a pressing need for improvement. We hope that his statement "To maximize the benefit to society, you need to not just do research but do it well" is a driver for all researchers.

5.1 | Education for statistics in practice

Following critical articles stating that much that is published in medical journals is statistically poor or even wrong (Altman, 1982; Altman et al., 1983), the *British Medical Journal*, asked Doug to contribute a series of short pieces on

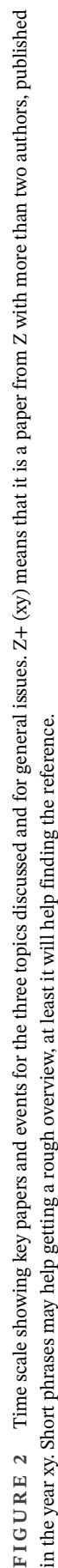


FIGURE 2 Time scale showing key papers and events for the three topics discussed and for general issues. Z+ (xy) means that it is a paper from Z with more than two authors, published in the year xy. Short phrases may help getting a rough overview, at least it will help finding the reference.

statistics and research for the journal. These were described as “fillers,” which would occupy spare parts of pages (of the then paper journal) at the end of research articles. They would not be refereed but Doug would take responsibility for them. He thought that this was a great opportunity to educate researchers with lower levels of statistical training, experience, and interest in basic statistical issues. To help protect him from mistakes (!), he suggested Martin Bland as a coauthor. They had already published 15 statistical articles and letters together, including two citation classics on agreement between methods of measurement (Altman & Bland, 1983; Bland & Altman, 1986; see Section 2.6).

The first Statistics Note was published in 1994 (Bland & Altman, 1994). In it, randomly generated numbers with five pairs of observations on each of five “subjects” were analyzed in two ways: using subject means for each variable and mixing observations from different “subjects” indiscriminately. The analysis of subject means found no significant correlation, as we would expect, there being no underlying relationship. The analysis of 25 observations produced a correlation coefficient closer to zero, but statistically significant. The incorrect nature of the analysis was both illustrated and explained. This was followed by eight more Statistics Notes in 1994, on regression toward the mean, diagnostic tests, one- and two-sided tests of significance, centiles and other quantiles, and matching. They continued to appear fitfully until Doug’s death in 2018. Doug and Martin agreed that other authors could be included, but that either Doug or Martin would always be an author. The other of the pair would check the Note and take editorial responsibility for it. Doug was much the more energetic collaborator and coauthored 13 Notes with eight colleagues, Martin only managing five Notes with one colleague. Doug was a natural collaborator.

The Notes began to be cited. According to Web of Science (WoS), at April 21, 2020, the first (Bland & Altman, 1994) had been cited 145 times. The most highly cited was an explanation of Cronbach’s alpha coefficient (Bland & Altman, 1997), with 2,053 citations on WoS. The second was on the Bonferroni method for multiple significance tests (Bland & Altman, 1995b), with 1,968 citations, and the third was on sensitivity and specificity of diagnostic tests (Altman & Bland, 1994), with 1,063 citations. The final Note, in 2018, was by Doug and Mohammad Mansournia, on population attributable fraction (Mansournia & Altman, 2018), cited 34 times. A full list of *BMJ* Statistics Notes, with links to their text in *BMJ*, can be found at <https://www-users.york.ac.uk/~mb55/pubs/pbstnote.htm>.

In addition to the *BMJ* series with Martin, Doug made sterling efforts to educate researchers in better use of statistical methods. Many summaries around specific statistical issues are given in the “explanation and elaboration” papers associated with reporting guidelines (see the third paragraphs in 3.2 [REMARK], 3.3 [TRIPOD], and in other E&E papers). Among many other relevant papers, we refer the reader to one that Doug and six eminent colleagues wrote about statistical tests, *p*-values, confidence intervals, and power, all key issues in statistical methodology (Greenland et al., 2016).

5.2 | Reporting of prognosis research

With his longstanding interests in evaluating and improving scientific reporting and in statistical considerations in studies of prognosis, Doug was naturally interested in the scientific reporting of prognosis research. This was emphasized by his involvement in the REMARK and TRIPOD reporting guideline initiatives to improve prognosis research. Doug’s motivation to improve reporting of prognosis research was to reduce waste from incomplete or unusable reports (Glasziou et al., 2014). Inadequate or incomplete reporting limits an assessment of a study’s methods and findings, bringing into question the usability of the research. This theme was present in much of Doug’s published work.

It is a long way from the first single studies to an evidence-based assessment of the prognostic value of a marker. In 2006, Doug and colleagues stressed the key role of IPD meta-analysis and stated that registration of prognostic marker studies would help to reduce publication bias (Sauerbrei, Holländer, Riley, & Altman, 2006). In a subsequent paper Doug, Richard, and Willi discussed the consequences of poor reporting of prognostic factor studies, including highlighting deficiencies in statistical analysis (e.g., dichotomizing continuous markers) (Riley, Sauerbrei, & Altman, 2009). With colleagues, Doug also wrote about the problems of selective reporting (Rifai, Altman, & Bossuyt, 2008), publication bias, and across-study heterogeneity, which are rife in prognosis studies and prevent formal evidence synthesis (Altman & Riley, 2005). In systematic reviews and meta-analyses, Yavchitz et al. (2016) identified 39 types of spin, which they classify and rank according to the severity. In 2018, Doug and his colleagues at the Centre for Statistics in Medicine evaluated reporting and levels of overinterpretation of prognostic factors in oncology (Kempf et al., 2018). They identified misleading reporting strategies used by authors that could influence how readers interpret study findings. Doug published extensively on the reporting and methodological conduct of prognostic factor research, both highlighting the problems of poor reporting and providing guidance for researchers to do better research (Riley, Moons et al., 2019).

In parallel to his interests in studies of single prognostic markers, Doug was also interested in prognostic models—using multiple prognostic covariates to make individualized predictions. He was involved in numerous systematic reviews evaluating the methodological conduct and reporting of prognostic models (Altman, 2009; Bouwmeester et al., 2012; Collins et al., 2014; Mallett et al., 2010). These reviews and many others have highlighted the problems that result from poor reporting. For example, an individualized prediction of prognosis requires a fully reported model. However, these reviews have all highlighted that models are often incompletely reported, so cannot be used. The resulting research waste could have been avoided by proper reporting.

An interesting consequence of the failure to report a prognostic model arose when Gary and Doug externally validated the QFracture score. They sought to independently evaluate the predictive accuracy of QFracture, a model to predict 10-year risk of osteoporotic fracture of any of several bones, including the hip, and compare it to another leading model in this area, FRAX (Collins & Altman, 2011). At the time, both models were being considered for inclusion in clinical guidelines. As Doug and Gary wrote in their resulting paper: *“The details and source code for the calculation of an individual’s risk using FRAX has to date not been published or released for independent evaluation. We sought to compare QFractureScores against FRAX for the hip fracture endpoint but our requests for an independent head to head comparison were not taken up by the developers of FRAX.”* In this instance, the lack of reporting of the model meant that the two leading models in this clinical area could not be directly compared, complicating the decision of which model to use.

In addition to reporting guidelines, Doug encouraged registration of prognosis studies (such as on clinicaltrials.gov) and publishing protocols to reduce selective reporting, improve transparency, and promote data sharing (Altman, 2014; Andre et al., 2011; Peat et al., 2014; Riley et al., 2009). In response, a journal called *Diagnostic and Prognostic Research* (diagnoprognres.biomedcentral.com) was established with Gary as editor-in-chief and Doug, Richard, Willi, and many others on the editorial board. The journal regularly publishes protocols for all types of prognosis research, including primary studies of prognostic factors and models, evidence synthesis, and methodological studies. Doug played an important role in early discussions with Gary to establish this journal.

5.3 | Structured reporting and study registration

Many papers have provided empirical evidence of bad reporting (Heus et al., 2018; Mallett et al., 2010; Sekula et al., 2017). It is evident that simply developing reporting guidelines and providing detailed background information is not enough to solve the problem. E&E papers are (too) long and do not seem to reach most of the relevant researchers. To alleviate the situation, Sauerbrei, Taube, McShane, Cavenagh, and Altman (2018) produced a considerably abridged version of the detailed REMARK E&E paper. The paper aims to serve as a brief guide to key issues for investigators planning tumor marker prognostic studies, and can be easily translated into other languages. Translations will help to bring key information to a larger global audience. Sauerbrei et al. (2018) emphasized that several cancer journals ask authors to follow the REMARK recommendations in their instructions to authors, and encouraged more journals to follow this example. Action by authors, reviewers, and editors is needed to improve reporting.

In 2014, Doug and colleagues published a book giving an overview of why reporting guidelines are needed and summarizing some of the major reporting guidelines, *Guidelines for Reporting Health Research – a user’s manual* (Moher et al., 2014). Doug was a coauthor of 12 of the 29 chapters. A list of “creator’s preferred bits” is included for each summarized guideline. For REMARK, Altman, McShane, Sauerbrei, Taube, and Cavenagh (2014) emphasized that estimates with confidence intervals need to be given, regardless of statistical significance of the estimate. They also stressed the importance of structured reporting, as proposed in the two-part REMARK profile.

The upper part of the profile gives details about the patient population, how the marker was handled in the analysis, which other variables were available, and key information about the outcome. The lower part of the profile details each analysis performed, including the number of patients and events and the amount of data available for each analysis. Although subgroup analyses are often conducted, relevant information, such as the effective sample size, is often omitted. Defining all details of the analysis part prospectively when designing a study would correspond to a very detailed statistical analysis plan (SAP). Guidelines for the content of SAPs have been proposed for clinical trials (Gamble et al., 2017). An SAP may have to be modified, for example, if important assumptions are violated. Any changes could also be noted in the paper’s REMARK profile. Readers would then see all analyses and would be able to distinguish between preplanned analyses and data-dependent modifications or additional analyses.

The REMARK E&E paper (Altman et al., 2012) followed Doug's principle that *"everything needs to be as simple as possible,"* presenting just two simple profiles that illustrate the key issues. Winzer, Buchholz, Schumacher, and Sauerbrei (2016) extended the profile by proposing additional information about the initial data analysis (Huebner, Vach, le Cessie, Schmidt, & Lusa, 2020) and checking relevant assumptions. Similar profiles have recently been developed for other types of studies.

In an editorial on a paper providing empirical evidence of the seriousness of selective reporting biases in cancer prognostic factor studies (Kyzas, Loizou, & Ioannidis, 2005), McShane, Altman, and Sauerbrei (2005) stated that *"The number of cancer prognostic markers that have been validated as clinically useful is pitifully small, despite decades of effort and money invested in marker research. For nearly all markers, the product has been a collection of studies that are difficult to interpret because of inconsistencies in conclusions or a lack of comparability. Small, underpowered studies; poor study design; varying and sometimes inappropriate statistical analyses; and differences in assay methods or endpoint definitions are but a few of the explanations that have been offered for this disappointing state of affairs."* Further empirical evidence of selective reporting biases was provided by Kyzas, Denaxa-Kyza, and Ioannidis (2007). They evaluated about 2,000 articles on cancer prognostic markers and showed that almost all of the articles reported statistically significant results. Publication bias is clearly an important issue in marker research.

Publication bias and hidden multiple-hypothesis testing distort the assessment of the true value of markers. Publication bias from preferential reporting of "positive" findings is well recognized. Hidden multihypothesis testing arises when several biomarkers are tested by different teams using the same samples. The more hypotheses (i.e., biomarker association with outcome) that are tested, the greater the risk of false-positive findings. These biases inflate the potential clinical validity and utility of published biomarkers, while negative results often remain hidden. The REMARK profile aims to facilitate more complete and better reporting of all analyses. Better reporting, the use of trial registries, and preregistration of all phase II and III trials will substantially improve prognosis research (Altman, 2014; Andre et al., 2011; Peat et al., 2014; Riley et al., 2009).

5.4 | Standardization and guidance for analysis

With the series in the BMJ, Doug and Martin Bland (see Section 5.1) aimed to provide basic information on statistical issues to researchers with limited knowledge in statistical methodology. However, weakness in many analyses clearly showed that many experienced statisticians also need guidance and that more standardization is required.

Doug emphasized the importance of these issues earlier than most other methodological researchers, as evidenced by his early role in groups developing reporting guideline and other initiatives. Work on REMARK (see Section 3.2) and COSMIC group discussions (see Section 2.5) showed that guidance for analysis was urgently needed. In 2011, Willi was mandated by the Epidemiology Subcommittee (EpidSC) of ISCB to start a project to create guidance for the analyses of observational data. Doug was a strong supporter of the project from the beginning, discussing the project's motivation, mission, structure, and main aims with Willi. The result of discussions among many researchers, including all COSMIC members, was the launch of the STRATOS initiative at ISCB 2013. Doug was the only speaker at the first STRATOS mini-symposium who gave two presentations (a general talk and a talk on behalf of the topic group TG5 "Study design"). He was one of five authors of a paper introducing STRATOS (Sauerbrei, Abrahamowicz, Altman, le Cessie, & Carpenter, 2014), which stated: *"The over-arching objective is to provide accessible and accurate guidance for data analysts with different levels of statistical education and interests, taking into account the differences in their training and experience."* Twenty years earlier, Doug stressed in his "scandal" article (Altman, 1994) that: *"Statistical refereeing is a form of firefighting. The time spent refereeing medical papers (often for little or no reward) would be much better spent in education and in direct participation in research as a member of the research team. There is, though, a serious shortage of statisticians to teach and, especially, to participate in research."*

Doug was a member of the STRATOS Steering Group and TG5, and cochaired with Willi the Contact Organizations panel. The first TG5 paper appeared soon after his death (Gail et al., 2019). In a report introducing the Design topic group to members of the International Biometric Society, Gail and Cadarette wrote *"Doug Altman was an inspirational leader in TG5, and his contributions continue to influence our thinking"* (Gail & Cadarette, 2019).

Beside of a small part of Doug's CONSORT work on reporting guidelines, we hardly touched his contributions to clinical trials. Nowadays, guidance is available for many parts of trials but until recently several gaps existed. With colleagues, Doug proposed standards for the development of a core outcome set and for the content of SAPs (Gamble et al., 2017; Kirkham et al., 2017; Williamson et al., 2017). Such guidance is needed to support transparency and reproducibility.

6 | DOUG ALTMAN—A PERSONAL NOTE BY PATRICK ROYSTON

I first met Doug in late 1976 when he and the late Michael Healy interviewed me for my first job in medical statistics, at the then UK Medical Research Council Clinical Research Centre (CRC) in north-west London. I was fortunate enough to be appointed to the post, which largely involved providing a statistical consultancy service to researchers at the CRC. Doug had already worked there for about a year in the Division of Computing and Statistics.

We worked as consulting statisticians learning an enormous amount about the practicalities and pitfalls of developing and executing medical research projects. I believe much of Doug's burning desire to improve the quality of the many aspects of statistics in medical research arose from this experience of the research "coal-face." One thing that became increasingly apparent to us was how weak (or nonexistent) was the planning of medical research, particularly study design and analysis of the resulting data. Our first question to a researcher consulting us for advice, usually on how best to analyze their data, was almost invariably "What is the question you want to answer?" It was amazing how few would-be researchers had thought about that. Doug's way of posing such questions and eliciting the answers enabled him to win the confidence of medical researchers while being able to be critical of their work almost unobtrusively and directing them to better ways of doing the research.

His experience as a consulting statistician also inspired him to write his best-selling textbook *Practical Statistics for Medical Research* (PSMR). I can personally vouch that the task took him more than 10 years of intermittent but determined effort fitted around his "day job." I recall him discussing with me how hard it was to find appropriate real data sets to illustrate statistical methods and principles in a direct and approachable way. He was absolutely determined to use only real-data examples throughout the book, a distinctive feature of the work. In my view and that of many clinicians, PSMR remains one of the best and most approachable texts of its type on the market. The ideas and approaches in the book also inspired a short course with the same name, which still runs to this day (see <https://www.ucl.ac.uk/statistics/psmr>). He was an inspiring teacher. His determination to use real examples shows he was grounded in actual research and the many issues it can throw up rather than being the type of statistician who was interested only in numbers or abstract mathematics.

We both left the CRC for new jobs on the same day (February 4, 1988). Our interests had developed rather differently over the decade at the CRC. Doug had become more and more passionate about improving the quality of medical research. I was inspired more by developing methodological issues in medical statistics, always with a view to transparency, common sense, and practical usefulness. Therewith arose my strong and continuing interest in statistical software and algorithms to embody and make accessible new techniques that promote good statistical practice. Although our paths subsequently diverged in terms of different roles and different places, we continued to collaborate successfully (and enjoyably) on matters of mutual interest essentially until his untimely death in 2018.

On any topic he was always able to see the big picture underlying a detailed methodological research effort. That was hugely helpful to a researcher like me.

On a personal level, Doug was a delight to have as a friend. He and I spoke, corresponded with and most importantly, met in person many times over our working and private lives. He had an enormous sense of fun and a spirit-lifting, infectious laugh. He had this characteristic way of throwing his head back as he laughed which made you want to join in, even if it took you a second or two to catch up and realize what his quick brain had spotted was the funny point of what you were discussing. He was amused and amusing without trivializing important questions. Yet, he could complain vocally and bitterly when something rather minor was not going his way. (I think he felt assaulted in the *amour propre*.)

One quirk that accorded beautifully with our research interests was his struggle with the cutpoint issue in daily life. For example, Doug loved photography and good-quality cameras. Suppose he was considering buying a new camera and had to decide how much to spend. What upper limit should he place on the price he was willing to pay? Dearer usually means better, but why not go even dearer (or cheaper)? Or even (seemingly) simpler still, when in a supermarket, which wine to choose? I do not think he ever really overcame the torment of such cutpoint placement issues. His family acknowledged his ability to look at lots of restaurants before returning to the first he had seen to actually eat!

In conclusion, it was a terrible loss for a huge number of people, including me, when Doug died from advanced colorectal cancer at the relatively young age of 69. He would undoubtedly have continued to contribute in a significant way to the quality and effectiveness of medical research, ultimately to the benefit of patients. On a human level, he was terrific fun to spend time with, and I miss him sorely.

7 | SUMMARY

Doug was the one of the first to review the statistical methods used in medical journals. Although reviewing statistical methodology in peer review is standard today, his 1982 overview and 1991 review of developments in the previous decade were lone voices (Altman, 1982, 1991c). Today, systematic reviews are considered a key task for researchers and overviews are usually conducted by cooperative groups, but Doug did this hard work on his own. It seems that many of his colleagues did not recognize the importance of this work or did not see it as something that could be funded.

As he so clearly expressed in his scandal article, Doug was convinced that the use of suitable statistical methods was a key issue in improving medical research (Altman, 1994). However, his early experience of statistical review for medical journals was negative, as he found “..limited evidence of its effectiveness ...” (Altman, 1998). Assessing the statistical reviewing policies of medical journals, Goodman, Altman, and George (1998) concluded “*Except in the largest circulation medical journals, the probability of formal methodological review of original research is fairly low. As readers and researchers depend on the journals to assess the validity of the statistical methods and logic used in published reports, this is potentially a serious problem.*”

Doug became deeply involved in the new movement for reporting guidelines and was one of its leading researchers when this topic was still ignored by most researchers and, to the best of our knowledge, the ISCB. The first invited ISCB session on this topic, entitled “Unbiased reporting, integrity and ethics,” was in 2015 in Utrecht, with Willi as a co-organizer and Doug, Patrick Bossuyt, and Lisa McShane as speakers. Today, it is becoming more accepted that “*Reporting research is as important a part of a study as its design or analysis*” (Jordan & Lewis, 2009). However, even excellent reporting does not help if a study is badly designed or analyzed. In response, the STRATOS initiative (Sauerbrei et al., 2014) aims to provide accessible and accurate guidance for relevant topics in the design and analysis of observational studies.

In 2000, Doug wrote about areas of medical statistics that had gained prominence in the 1990s (Altman, 2000). Unfortunately, his broad observations would still be true today:

- (i) The misuse of statistics is very important.
- (ii) A general climate of sloppiness is bad for science.
- (iii) Statistics is much more subjective (and difficult) than is usually acknowledged (this is why statisticians have not been replaced by computers).
- (iv) Major improvements in the quality of research published in medical journals are unlikely in the present research climate.
- (v) Too much research is done primarily to benefit the careers of researchers.
- (vi) It need not be like this!

No, it need not be like this! Since the publication of these statements, Doug has inspired many colleagues and younger students with his visionary thoughts. Many have followed in his steps to improve the use of statistics in medical research. We hope that this article inspires you to follow Doug’s ideas to improve the quality of methodological and medical research.

ACKNOWLEDGMENTS

We thank Jannik Braun and Andreas Ott from the Institute of Medical Biometry and Statistics, Medical Center, University of Freiburg, for administrative assistance and Dr Jennifer A de Beyer of the Centre for Statistics in Medicine, University of Oxford, for English language editing. We thank the reviewer Ewout Steyerberg to propose the time scale shown in Figure 2 with relevant events and publications.

ORCID

Willi Sauerbrei  <https://orcid.org/0000-0002-6792-4123>

Martin Bland  <https://orcid.org/0000-0002-9525-5334>

Stephen J. W. Evans  <https://orcid.org/0000-0002-1474-2596>

Martin Schumacher  <https://orcid.org/0000-0002-5037-0099>

Gary S. Collins  <https://orcid.org/0000-0002-2772-2316>

REFERENCES

- Altman, D. G. (1982). Statistics in medical journals. *Statistics in Medicine*, 1, 59–71.
- Altman, D. G. (1991a). *Practical statistics for medical research*. London: Chapman & Hall.
- Altman, D. G. (1991b). Categorising continuous variables. *British Journal of Cancer*, 64, 975.
- Altman, D. G. (1991c). Statistics in medical journals: Developments in the 1980s. *Statistics in Medicine*, 10(12), 1897–1913.
- Altman, D. G. (1993). Construction of age-related reference centiles using absolute residuals. *Statistics in Medicine*, 12(10), 917–924.
- Altman, D. G. (1994). The scandal of poor medical research. *BMJ (Clinical Research Edition)*, 308(6924), 283–284.
- Altman, D. G. (1998). Statistical reviewing for medical journals. *Statistics in Medicine*, 17, 2661–2674.
- Altman, D. G. (2000). Statistics in medical journals: Some recent trends. *Statistics in Medicine*, 19(23), 3275–3289.
- Altman, D. G. (2001). Systematic reviews of evaluations of prognostic variables. *BMJ (Clinical Research Edition)*, 323(7306), 224–228.
- Altman, D. G. (2009). Prognostic models: A methodological framework and review of models for breast cancer. *Cancer Investigation*, 27, 235–243.
- Altman, D. G. (2014). The time has come to register diagnostic and prognostic research. *Clinical Chemistry*, 60, 580–582.
- Altman, D. G., & Andersen, P. K. (1989). Bootstrap investigation of the stability of a Cox regression model. *Statistics in Medicine*, 8, 771–783.
- Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: The analysis of method comparison studies. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 32(3), 307–317.
- Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*, 308(6943), 1552.
- Altman, D. G., & Chitty, L. S. (1994). Charts of fetal size: 1. Methodology. *BJOG*, 101(1), 29–34.
- Altman, D. G., Gore, S. M., Gardner, M. J., & Pocock, S. J. (1983). Statistical guidelines for contributors to medical journals. *British Medical Journal*, 286(6376), 1489–1493.
- Altman, D. G., Lausen, B., Sauerbrei, W., & Schumacher, M. (1994). Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *Journal of the National Cancer Institute*, 86(11), 829–835.
- Altman, D. G., & Lyman, G. H. (1998). Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Research and Treatment*, 52, 289–303.
- Altman, D. G., Machin, D., Bryant, T., & Gardner, M. (2000). *Statistics with confidence: Confidence intervals and statistical guidelines*. BMJ Books.
- Altman, D. G., McShane, L., Sauerbrei, W., & Taube, S. E. (2012). Reporting recommendations for tumor marker prognostic studies (REMARK): Explanation and elaboration. *PLoS Medicine*, 9(5), E1001216.
- Altman, D. G., McShane, L. M., Sauerbrei, W., Taube, S. E., & Cavenagh, M. M. (2014). REMARK (Reporting Recommendations for Tumor MARKer Prognostic Studies). In D. Moher, D. G. Altman, K. F. Schulz, I. Simera, & E. Wager, (Eds.). *Guidelines for reporting health research: A user's manual* (pp. 241–249). John Wiley & Sons, Ltd.
- Altman, D. G., & Riley, R. D. (2005). Primer: An evidence-based approach to prognostic markers. *Nature Clinical Practice Oncology*, 2, 466–472.
- Altman, D. G., & Royston, P. (2000). What do we mean by validating a prognostic model? *Statistics in Medicine*, 19, 453–473.
- Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ (Clinical Research Edition)*, 332(7549), 1080.
- Altman, D. G., Trivella, M., Pezzella, F., Harris, A. L., & Pastorino, U. (2006). Systematic review of multiple studies of prognosis: The feasibility of obtaining individual patient data. In J. L. Auget, N. Balakrishnan, M. Mesbah, & G. Molenberghs, (Eds.). *Advances in statistical methods for the health sciences* (pp. 3–18). Boston: Birkhäuser.
- Andre, F., McShane, L. M., Michiels, S., Ransohoff, D. F., Altman, D. G., Reis-Filho, J. S., ... Lajos Pusztai, L. (2011). Biomarker studies: A call for a comprehensive biomarker study registry. *Nature Reviews Clinical Oncology*, 8(3), 171–176.
- Andrew, E., Anis, A., Chalmers, T., Cho, M., Clarke, M., Felson, D., ... Wells, G. (1994). A proposal for structured reporting of randomized controlled trials. *JAMA*, 272(24), 1926–1931.
- Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 327(8476), 307–310.
- Bland, J. M., & Altman, D. G. (1977). Enteric disease in San Francisco. *Lancet*, 310(8032), 306.
- Bland, J. M., & Altman, D. G. (1992). This week's citation classic: Comparing methods of clinical measurement. *Current Contents*, CM20(40), 8.
- Bland, J. M., & Altman, D. G. (1994). Correlation, regression, and repeated data. *BMJ: British Medical Journal*, 308(6933), 896.
- Bland, J. M., & Altman, D. G. (1995a). Comparing two methods of clinical measurement: A personal history. *International Journal of Epidemiology*, 24(Supplement_1), S7–S14.
- Bland, J. M., & Altman, D. G. (1995b). Statistics notes. Multiple significance tests: The Bonferroni method. *BMJ: British Medical Journal*, 310, 170.
- Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ: British Medical Journal*, 314(7080), 572.
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135–160.
- Boutron, I., Dutton, S., Ravaud, P., & Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*, 303(20), 2058–2064.
- Bouwmeester, W., Zuithoff, N. P., Mallett, S., Geerlings, M. I., Vergouwe, Y., Steyerberg, E. W., ... Moons, K. G. (2012). Reporting and methods in clinical prediction research: A systematic review. *PLoS Medicine*, 9(5), e1001221.
- Burton, A., & Altman, D. G. (2004). Missing covariate data within cancer prognostic studies: A review of current reporting and proposed guidelines. *British Journal of Cancer*, 91(1), 4–8.
- Campbell, M. K., Elbourne, D. R., & Altman, D. G. (2004). CONSORT statement: Extension to cluster randomised trials. *BMJ (Clinical Research Edition)*, 328(7441), 702–708.

- Chen, C. H., & George, S. L. (1985). The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Statistics in Medicine*, 4(1), 39–46.
- Christensen, E., Neuberger, J., Crowe, J., Altman, D. G., Popper, H., Portmann, B., ... Williams, R. (1985). Beneficial effect of azathioprine and prediction of prognosis in primary biliary cirrhosis: Final results of an international trial. *Gastroenterology*, 89(5), 1084–1091.
- Clarke, B. F., & Campbell, I. W. (1975). Long-term comparative trial of glibenclamide and chlorpropamide in diet-failed, maturity-onset diabetics. *Lancet*, 305(7901), 246–248.
- Collins, G. S. (2019). Doug Altman: A tribute. *ISCB News*.
- Collins, G. S., & Altman, D. G. (2011). Predicting risk of osteoporotic and hip fracture in the United Kingdom: Prospective independent and external validation of QFractureScores. *BMJ (Clinical Research Edition)*, 342, i3651.
- Collins, G. S., de Groot, J. A., Dutton, S., Omar, O., Shanyinde, M., Tajar, A., ... Altman, D. G. (2014). External validation of multivariable prediction models: A systematic review of methodological conduct and reporting. *BMC Medical Research Methodology*, 14(1), 40.
- Collins, G. S., & Moons, K. G. (2019). Reporting of artificial intelligence prediction models. *Lancet*, 393(10181), 1577–1579.
- Collins, G. S., Ogundimu, E. O., Cook, J. A., Manach, Y. L., & Altman, D. G. (2016). Sample size considerations for the external validation of a multivariable prognostic model: A resampling study. *Statistics in Medicine*, 35(2), 214–226.
- Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Annals of Internal Medicine*, 162(1), 55–63.
- Debray, T. P., Damen, J. A., Snell, K. I., Ensor, J., Hooft, L., Reitsma, J. B., ... Moons, K. G. (2017). A guide to systematic review and meta-analysis of prediction model performance. *BMJ (Clinical Research Edition)*, 356, i6460.
- Deeks, J. J., Hopewell, S., Moher, D., Higgins, J. P. T., Moons, K. G. M., Chandler, J., & Antes, G. (2018). Doug Altman's legacy to Cochrane and evidence synthesis. *Cochrane Database of Systematic Reviews*, 8(6933). <https://doi.org/10.1002/14651858.ED000127>.
- Dritz, S., Ainsworth, T., Back, A., Boucher, L., Garrard, W., Palmer, R., & River, E. (1977). Patterns of sexually transmitted enteric diseases in a city. *Lancet*, 310(8027), 3–4.
- Evans, S. (2018). Obituary: Doug Altman. *Biometric Bulletin*, 35(3), 4.
- Gail, M. H., Altman, D. G., Cadarette, S. M., Collins, G., Evans, S. J., Sekula, P., ... Woodward, M. (2019). Design choices for observational studies of the effect of exposure on disease incidence. *BMJ Open*, 9(12), e31031.
- Gail, M. H., & Cadarette, S. on behalf of STRATOS TG5. (2019). STREngthening Analytical Thinking for Observational Studies (STRATOS): Introducing the Study Design Topic Group (TG5). *Biometric Bulletin*, 36(2), 12–13.
- Gamble, C., Krishan, A., Stocken, D., Lewis, S., Juszczak, E., Doré, C., ... Loder, E. (2017). Guidelines for the content of statistical analysis plans in clinical trials. *JAMA*, 318(23), 2337–2343.
- Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., ... Wager, E. (2014). Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*, 383(9913), 267–276.
- Goodman, S. N., Altman, D. G., & George, S. L. (1998). Statistical reviewing policies of medical journals. *Journal of General Internal Medicine*, 13, 753–756.
- Gore, S. M., & Altman, D. G. (1982). *Statistics in practice*. London: British Medical Association.
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, 31, 337–350.
- Hall, P. A., & Going, J. J. (1999). Predicting the future: A critical appraisal of cancer prognosis studies. *Histopathology*, 35, 489–494.
- Hayes, D. F., Bast, R. C., Desch, C. E., Fritsche, H. Jr., Kemeny, N. E., Jessup, J. M., ... Winn, R. J. (1996). Tumor marker utility grading system: A framework to evaluate clinical utility of tumor markers. *Journal of the National Cancer Institute*, 88, 1456–1466.
- Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal*, 60(3), 431–449.
- Hemingway, H., Croft, P., Perel, P., Hayden, J. A., Abrams, K., Timmis, A., ... Riley, R. D.; for the PROGRESS Group. (2013). Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. *BMJ (Clinical Research Edition)*, 346, e5595.
- Heus, P., Damen, J. A. A. G., Pajouheshnia, R., Scholten, R. J. P. M., Reitsma, J. B., Collins, G. S., ... Hooft, L. (2018). Poor reporting of multivariable prediction model studies: Towards a targeted implementation strategy of the TRIPOD statement. *BMC Medicine*, 16, 120.
- Heus, P., Damen, J. A. A. G., Pajouheshnia, R., Scholten, R. J., Reitsma, J. B., Collins, G. S., ... Hooft, L. (2019). Uniformity in measuring adherence to reporting guidelines: The example of TRIPOD for assessing completeness of reporting of prediction model studies. *BMJ Open*, 9(4), e025611.
- Hingorani, A. D., van der Windt, D. A., Riley, R. D., Abrams, K., Moons, K. G., Steyerberg, E. W., ... Hemingway, H. (2013). Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ (Clinical Research Edition)*, 346, e5793.
- Huebner, M., Vach, W., le Cessie, S., Schmidt, C. O., & Lusa, L. (2020). Hidden analyses: A review of reporting practice and recommendations for more transparent reporting of initial data analyses. *BMC Medical Research Methodology*, 20(1), 1–10.
- Ioannidis, J. P. A., Evans, S. J. W., Gøtzsche, P. C., O'Neill, R. T., Altman, D. G., Schulz, K., & Moher, D. (2004). Better reporting of harms in randomized trials: An extension of the CONSORT statement. *Annals of Internal Medicine*, 141(10), 781–788.
- Jordan, K. P., & Lewis, M. (2009). Improving the quality of reporting of research studies. *Musculoskeletal Care*, 7, 137–142.
- Kempf, E., de Beyer, J. A., Cook, J., Holmes, J., Mohammed, S., Nguyễn, L., ... Collins, G. S. (2018). Overinterpretation and misreporting of prognostic factor studies in oncology: A systematic review. *British Journal of Cancer*, 119(10), 1288–1296.
- Kirkham, J. J., Davis, K., Altman, D. G., Blazeby, J. M., Clarke, M., Tunis, S., & Williamson, P. R. (2017). Core outcome Set-STAndards for development: The COS-STAD recommendations. *PLoS Medicine*, 14(11), e1002447.

- Kleinert, S., & Horton, R. (2014). How should medical science change? *Lancet*, 383(9913), 197–198.
- Kyzas, P. A., Loizou, K. T., & Ioannidis, J. P. A. (2005). Selective reporting biases in cancer prognostic factor studies. *Journal of the National Cancer Institute*, 97, 1043–1055.
- Kyzas, P. A., Denaxa-Kyza, D., & Ioannidis, J. P. (2007). Almost all articles on cancer prognostic markers report statistically significant results. *European Journal of Cancer*, 43(17), 2559–2579.
- Lausen, B., & Schumacher, M. (1992). Maximally selected rank statistics. *Biometrics*, 48(1), 73–85.
- Mallett, S., Royston, P., Dutton, S., Waters, R., & Altman, D. G. (2010). Reporting methods in studies developing prognostic models in cancer: A review. *BMC Medicine*, 8(1), 20.
- Mallett, S., Royston, P., Waters, R., Dutton, S., & Altman, D. G. (2010). Reporting performance of prognostic models in cancer: A review. *BMC Medicine*, 8(1), 21.
- Mallett, S., Timmer, A., Sauerbrei, W., & Altman, D. G. (2010). Reporting of prognostic studies of tumour markers: A review of published articles in relation to REMARK guidelines. *British Journal of Cancer*, 102(1), 173–180.
- Mansournia, M. A., & Altman, D. G. (2018). Population attributable fraction. *BMJ (Clinical Research Edition)*, 360, k757.
- Matthews, R., Chalmers, I., & Rothwell, P. (2018). Douglas G Altman: Statistician, researcher, and driving force behind global initiatives to improve the reliability of health research. *BMJ (Clinical Research Edition)*, 361, k2588.
- McShane, L. M., Aamodt, R., Cordon-Cardo, C., Cote, R., Faraggi, D., Fradet, Y., ... Waldman, F. M., the National Cancer Institute Bladder Tumor Marker Network. (2000). Reproducibility of p53 immunohistochemistry in bladder tumors. *Clinical Cancer Research*, 6, 1854–1864.
- McShane, L. M., Altman, D. G., & Sauerbrei, W. (2005). Identification of clinically useful cancer prognostic factors: What are we missing? (Editorial). *Journal of the National Cancer Institute*, 97, 1023–1025.
- McShane, L. M., Altman, D. G., Sauerbrei, W., Taube, S. E., Gion, M., & Clark, G. M., for the Statistics Subcommittee of the NCI-EORTC Working on Cancer Diagnostics. (2005). REporting recommendations for tumor MARKer prognostic studies (REMARK). *Journal of the National Cancer Institute*, 97, 1180–1184.
- Miller, R., & Siegmund, D. (1982). Maximally selected chi square statistics. *Biometrics*, 38(4), 1011–1016.
- Moher, D., Altman, D. G., Schulz, K. F., Simera, I., & Wager, E. (2014). *Guidelines for reporting health research: A user's manual*. John Wiley & Sons, Ltd.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097.
- Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*, 357(9263), 1191–1194.
- Moher, D., Schulz, K. F., Simera, I., & Altman, D. G. (2010). Guidance for developers of health research reporting guidelines. *PLoS Medicine*, 7(2), e100217.
- Moons, K. G., Altman, D. G., Reitsma, J. B., Ioannidis, J. P., Macaskill, P., Steyerberg, E. W., ... Collins, G. S. (2015). Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine*, 162(1), W1–W73.
- Ogundimu, E. O., Altman, D. G., & Collins, G. S. (2016). Adequate sample size for developing prediction models is not simply related to events per variable. *Journal of Clinical Epidemiology*, 76, 175–182.
- Ohuma, E. O., & Altman, D. G., International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st Project). (2019). Statistical methodology for constructing gestational age-related charts using cross-sectional and longitudinal data: The INTERGROWTH-21st project as a case study. *Statistics in Medicine*, 38(19), 3507–3526.
- Page, M. J., Altman, D. G., McKenzie, J. E., Shamseer, L., Ahmadzai, N., Wolfe, D., ... Moher, D. (2018). Flaws in the application and interpretation of statistical analyses in systematic reviews of therapeutic interventions were common: A cross-sectional analysis. *Journal of Clinical Epidemiology*, 95, 7–18.
- Papageorgiou, A. T., Kennedy, S. H., Salomon, L. J., Altman, D. G., Ohuma, E. O., Stones, W., ... Villar, J. (2018). The INTERGROWTH-21st fetal growth standards: Toward the global integration of pregnancy and pediatric care. *American Journal of Obstetrics and Gynecology*, 218(2), S630–S640.
- Peat, G., Riley, R. D., Croft, P., Morley, K. I., Kyzas, P. A., Moons, K. G. M., ... Hemingway, H.; for the PROGRESS Group. (2014). Improving the transparency of prognosis research: The role of reporting, data sharing, registration, and protocols. *PLoS Medicine*, 11(7), e1001671.
- Piaggio, G., Elbourne, D. R., Altman, D. G., Pocock, S. J., & Evans, S. J. W. (2006). Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. *JAMA*, 295(10), 1152–1160.
- Pocock, S. J., Hughes, M. D., & Lee, R. J. (1987). Statistical problems in the reporting of clinical trials. A survey of three medical journals. *New England Journal of Medicine*, 317(7), 426–432.
- Rennie, D. (2019). A tribute to Douglas G. Altman, DSc. *JAMA*, 321(16), 1621.
- Rifai, N., Altman, D. G., & Bossuyt, P. (2008). Reporting bias in diagnostic and prognostic studies: Time for action. *Clinical Chemistry*, 54, 1101–1103.
- Riley, R. D., Hayden, J. A., Steyerberg, E. W., Moons, K. G., Abrams, K., Kyzas, P. A., ... Hemingway, H. (2013). Prognosis Research Strategy (PROGRESS) 2: Prognostic factor research. *PLoS Medicine*, 10(2), e1001380.
- Riley, R. D., Moons, K. G., Snell, K. I., Ensor, J., Hooft, L., Altman, D. G., ... Debray, T. P. (2019). A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ (Clinical Research Edition)*, 364, k4597.

- Riley, R. D., Ridley, G., Williams, K., Altman, D. G., Hayden, J., & De Vet, H. C. (2007). Prognosis research: Toward evidence-based results and a Cochrane methods group. *Journal of Clinical Epidemiology*, 60(8), 863–865.
- Riley, R. D., Sauerbrei, W., & Altman, D. G. (2009). Prognostic markers in cancer: The evolution of evidence from single studies to meta-analysis, and beyond. *British Journal of Cancer*, 100(8), 1219–1229.
- Riley, R. D., van der Windt, D., Croft, P., & Moons, K. G. M., (Eds.) (2019). *Prognosis research in healthcare: Concepts, methods, and impact*. Oxford University Press.
- Royston, P., & Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(3), 429–467.
- Royston, P., & Altman, D. G. (1997). Approximating statistical functions by using fractional polynomial regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 46(3), 411–422.
- Royston, P., & Altman, D. G. (2013). External validation of a Cox prognostic model: Principles and methods. *BMC Medical Research Methodology*, 13(1), 33.
- Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, 25(1), 127–141.
- Royston, P., & Sauerbrei, W. (2008). *A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables*. Wiley & Sons Ltd.
- Sauerbrei, W., Abrahamowicz, M., Altman, D. G., le Cessie, S., & Carpenter, J.; on behalf of the STRATOS initiative. (2014). STREngthening Analytical Thinking for Observational Studies: The STRATOS initiative. *Statistics in Medicine*, 33, 5413–5432.
- Sauerbrei, W., Holländer, N., Riley, R. D., & Altman, D. G. (2006). Evidence-based assessment and application of prognostic markers: The long way from single studies to meta-analysis. *Communications in Statistics - Theory and Methods*, 35(7), 1333–1342.
- Sauerbrei, W., & Royston, P. (1999). Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society, A*, 162, 71–94.
- Sauerbrei, W., & Schumacher, M. (1992). A bootstrap resampling procedure for model building: Application to the Cox regression model. *Statistics in Medicine*, 11, 2093–2109.
- Sauerbrei, W., Taube, S. E., McShane, L. M., Cavenagh, M. M., & Altman, D. G. (2018). Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): An abridged explanation and elaboration. *Journal of the National Cancer Institute*, 110(8), 803–811.
- Sekula, P., Mallett, S., Altman, D. G., & Sauerbrei, 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*, 12(6), e0178531.
- Sigurdsson, H., Baldetorp, B., Borg, Å., Dalberg, M., Fernö, M., Killander, D., ... Ranstam, J. (1990). Flow cytometry in primary breast cancer: Improving the prognostic value of the fraction of cells in the S-phase by optimal categorisation of cut-off levels. *British Journal of Cancer*, 62(5), 786.
- Simera, I., Moher, D., Hirst, A., Hoey, J., Schulz, K. F., & Altman, D. G. (2010). Transparent and accurate reporting increases reliability, utility, and impact of your research: Reporting guidelines and the EQUATOR Network. *BMC Medicine [Electronic Resource]*, 8, 24.
- Simon, R., & Altman, D. G. (1994). Statistical aspects of prognostic factor studies in oncology. *British Journal of Cancer*, 69(6), 979–985.
- Steyerberg, E. W., Moons, K. G., van der Windt, D. A., Hayden, J. A., Perel, P., Schroter, S., ... Altman, D. G., for the PROGRESS Group. (2013). Prognosis Research Strategy (PROGRESS) 3: Prognostic model research. *PLoS Medicine*, 10(2), e1001381.
- Trivella, M. (2019). *The Albus Dumbledore of the medical research world—What would Doug Altman think?* BMJ Opinion. Retrieved from <https://blogs.bmj.com/bmj/2019/09/25/marialena-trivella-the-albus-dumbledore-of-the-medical-research-world-what-would-doug-altman-think/>
- Trivella, M., Pezzella, F., Pastorino, U., Harris, A. L., & Altman, D. G., for the Prognosis In Lung Cancer (PILC) Collaborative Study Group. (2007). Microvessel density as a prognostic factor in non-small-cell lung carcinoma: A meta-analysis of individual patient data. *Lancet Oncology*, 8(6), 488–499.
- Van Noorden, R., Maher, B., & Nuzzo, R. (2014). The top 100 papers. *Nature News*, 514(7524), 550.
- van Smeden, M., de Groot, J. A., Moons, K. G., Collins, G. S., Altman, D. G., Eijkemans, M. J., & Reitsma, J. B. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Medical Research Methodology*, 16(1), 163.
- van Smeden, M., Moons, K. G., de Groot, J. A., Collins, G. S., Altman, D. G., Eijkemans, M. J., & Reitsma, J. B. (2019). Sample size for binary logistic prediction models: Beyond events per variable criteria. *Statistical Methods in Medical Research*, 28(8), 2455–2474.
- Villar, J., Ismail, L. C., Victora, C. G., Ohuma, E. O., Bertino, E., Altman, D. G., ... Kennedy, S. H. (2014). International standards for newborn weight, length, and head circumference by gestational age and sex: The Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet*, 384(9946), 857–868.
- von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for reporting observational studies. *Annals of Internal Medicine*, 147(8), 573–577.
- Watts, G. (2018). Douglas Graham Altman. *Lancet*, 392(10141), 24.
- Williamson, P. R., Altman, D. G., Bagley, H., Barnes, K. L., Blazeby, J. M., Brookes, S. T., ... Young, B. (2017). The COMET handbook: Version 1.0. *Trials*, 18(3), 280.
- Winzer, K. J., Buchholz, A., Schumacher, M., & Sauerbrei, W. (2016). Improving the prognostic ability through better use of standard clinical data — The Nottingham Prognostic Index as an example. *PLoS One*, 11(3), e0149977.

- Wyatt, J. C., & Altman, D. G. (1995). Commentary: Prognostic models: Clinically useful or quickly forgotten? *BMJ (Clinical Research Edition)*, 311, 1539.
- Yavchitz, A., Ravaud, P., Altman, D. G., Moher, D., Hrobjartsson, A., Lasserson, T., & Boutron, I. (2016). A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity. *Journal of Clinical Epidemiology*, 75, 56–65.
- Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D. G., Tunis, S., Haynes, B., ... Moher, D., CONSORT Group. (2008). Improving the reporting of pragmatic trials: An extension of the CONSORT statement. *BMJ (Clinical Research Edition)*, 337, a2390.

How to cite this article: Sauerbrei W, Bland M, Evans SJW, et al. Doug Altman: Driving critical appraisal and improvements in the quality of methodological and medical research. *Biometrical Journal*. 2020;1–21.

<https://doi.org/10.1002/bimj.202000053>