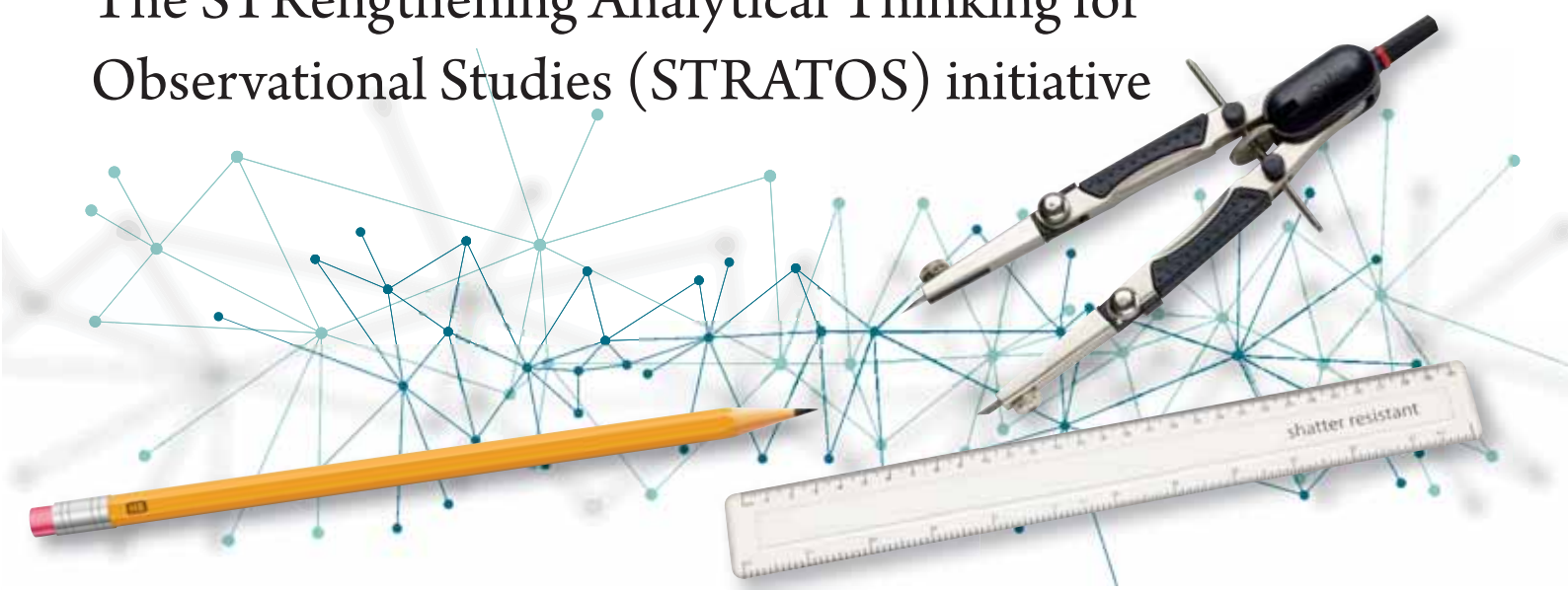
Guidance for designing and analysing observational studies:

The STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative



Willi Sauerbrei[1], **Gary S. Collins**[2], **Marianne Huebner**[3], **Stephen D. Walter**[4], **Suzanne M. Cadarette**[5], and **Michal Abrahamowicz**[6] on behalf of the **STRATOS initiative**

1 Faculty of Medicine and Medical Center, University of Freiburg, Germany
2 Centre for Statistics in Medicine, University of Oxford, UK
3 Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA
4 Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada
5 Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, Canada
6 Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada

Correspondence to:

Dr Willi Sauerbrei
Faculty of Medicine and Medical Center – University of Freiburg
Institute for Medical Biometry and Statistics
Stefan-Meier-Str. 26
79104 Freiburg
Germany
wfs@imbi.uni-freiburg.de
+49 761 203 6669

Abstract

Observational studies pose a number of biostatistical challenges. Methodological approaches have grown exponentially, but most are rarely applied in the real world. The STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative is an international collaboration that was formed to provide guidance to help bridge the gap between methodological innovation and application. STRATOS is focused on identifying issues and promising approaches for planning and analysing observational studies. Crucially, STRATOS will communicate its findings to a wide audience with different levels of statistical knowledge. In this article, we provide an example illustrating the need for such guidance and describe the structure, general approach, and general outlook of the STRATOS initiative.

Introduction

Substantial progress has been made in the methodology of clinical and epidemiological studies over the past few decades. However, research quality in the health sciences has not always matched this progress. Altman expressed several critical concerns in an editorial titled "The scandal of poor medical research",[1] and Ioannidis argued that most published research findings are false.[2] In 2014, *The Lancet* started a series called "Research: Increasing Value, Reducing Waste".[3] The question is no longer whether medical science needs to change but rather "How should medical science change?" [4] An estimated 85% of research investment is wasted.[5] A substantial part of this is due to weaknesses in the design, analysis, and reporting of medical research.[6] For studies on prognostic factors, Sauerbrei described several deficiencies and illustrated weaknesses and false conclusions that may arise from the use of inappropriate statistical methods in data analysis. [7]

Problems with the quality of medical research and the importance of using accurate statistical methodology are also discussed outside the medical literature. In the article "Unreliable research: Trouble at the lab",[8] the *Economist* summarised the current situation:

> *Scientists' grasp of statistics has not kept pace with the development of complex mathematical techniques for crunching data. Some scientists use inappropriate techniques because those are the ones they feel comfortable with; others latch on to new ones without understanding their subtleties. Some just rely on the methods built into their software, even if they don't understand them.*

Pointing to insufficient education of many researchers who attempt to use advanced statistical packages, Vickers[9] recently argued that

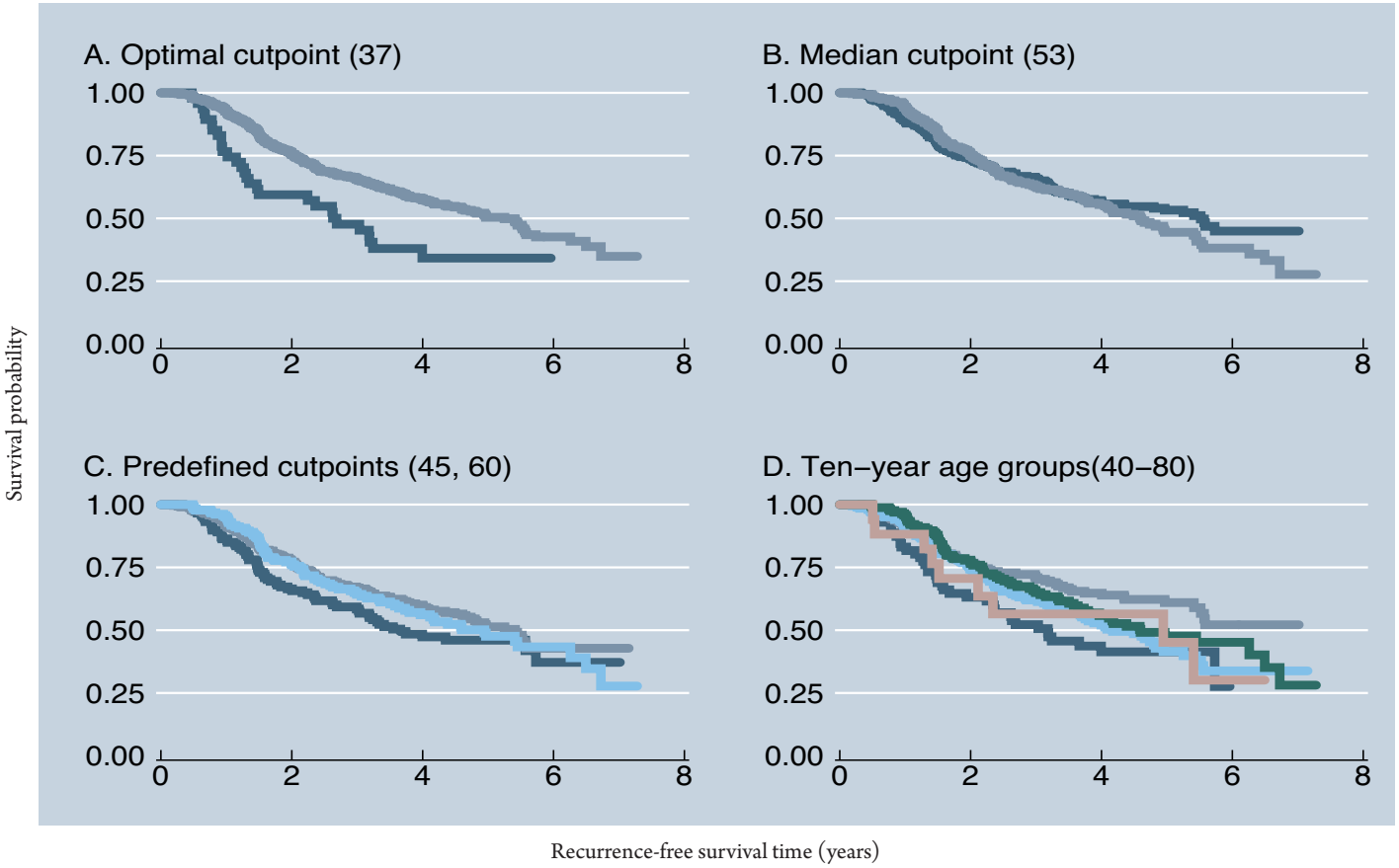


Figure 1. Kaplan-Meier plots using four alternative methods of categorizing age

(A) "Optimal" cutpoint (≤37 vs >37 years). (B) Cutpoint based on the median (≤53 vs >53 years). (C) Cutpoints predefined from earlier analyses and based on menopause (≤45, 46 to 60, >60 years); (D) 10-year increments (≤40; 41–50; 51–60; 61–70; >70). In all panels, the youngest age group is indicated in dark blue.

A mistake in the operating room can threaten the life of one patient; a mistake in statistical analysis or interpretation can lead to hundreds of early deaths. So it is perhaps odd that, while we allow a doctor to conduct surgery only after years of training, we give SPSS to almost anyone.

In this article, we present the STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative,[10] which is intended to develop guidance for planning and analysing observational studies. Below, we provide an example of difficulties in selecting an appropriate statistical method, which illustrates the need for such guidance.

Difficulties in selecting an appropriate statistical method: the example of handling one continuous variable

In medicine, continuous measurements, such as age, weight, and blood pressure, are often used to assess risk, predict outcome, or select a therapy. Background knowledge or the type of question can strongly influence how continuous variables are used, but the method for analysing continuous variables must often be selected.

Continuous variables are commonly assumed to be linearly related to outcome, but this is often inappropriate. To avoid this assumption, cutpoints are often applied to categorise the variable, implying regression models with step functions. At first glance, this seems to simplify analysis and aid interpretation, but categorising discards information and raises critical issues, such as how many cutpoints to use and where to place them. In addition, cutpoints create biologically implausible step functions, whereby individuals above and below a cutpoint have different risks – which is nonsensical.[1,11,12]

Consider the prognostic effect of age on recurrence-free survival (RFS) in breast cancer patients, an example discussed in detail in "Continuous variables: to categorise or to model" by Sauerbrei and Royston[13] and using data from a study by the German Breast Cancer Study Group. The data are publicly available and further details about the study have been published.[14] To analyse the impact of age on RFS, age categories can be set using various strategies. For this analysis, we present the four options: (1) an "optimal" cutpoint to create two groups; (2) the median as the cutpoint for two groups; (3) three groups based on a menopausal criterion; and (4) 10-year increments.

An "optimal" cutpoint of 37 years results in a large difference in survival curves between two groups: Younger patients have much lower RFS probabilities than older patients (Figure 1A). The corresponding hazard ratio estimate (Cox model) for older patients is 0.54 (95% confidence interval 0.37, 0.80). The difference in RFS between the two age groups disappears if the cutpoint is taken at the median (53 years) as indicated by a hazard ratio of 1.1 (95% confidence interval 0.88, 1.39) (Figure 1B). When age is divided into three groups according to predefined cutpoints of 45 and 60 years (premenopausal, mix, and postmenopausal), RFS differences again are small (Figure 1C). Finally, when ages are split into five 10-year age groups starting at 40 years, the probability of RFS appears slightly lower for patients under 40 years of age, with only negligible differences between the other groups, revealing that age is not linearly related to RFS (Figure 1D).

Thus, using cutpoints can lead to different and inconsistent results, even when only one

variable is considered. [1] An alternative and more appropriate approach is to estimate the functional form of a continuous variable on the outcome, for example using spline-based approaches or fractional polynomials. In contrast to cutpoint approaches, splines and fractional polynomials use the full information from a continuous variable and have several advantages.[10,11] In the breast cancer example, the fractional polynomial approach clearly showed that age has a strong nonlinear effect on RFS. For young patients (about 30 years of age), the relative risk of an event is high. The relative risk rapidly decreases with age, and for patients aged 40 or more years, age has a negligible influence on RFS.[13] Because there is no widely accepted agreement about how to handle continuous variables, many analysts proceed with cutpoint approaches. Indeed, introductory graduate-level courses often encourage this. Guidance that includes evidence of the advantages and disadvantages of competing strategies is thus needed.

The trickle-down effect of using cutpoints

Using multiple strategies for cutpoints in individual studies complicates assessing the risk or prognostic effect of a continuous variable in a meta-analysis. Altman et al. (1994) found 19 different cutpoints used in the literature to categorise S-phase fraction as a prognostic factor in breast cancer.[1] Conducting a meta-analyis to compare low vs. high S-phase fraction values could be done but cannot be interpreted because a patient with a specific S-phase fraction value could belong to the "low" group in one study and the "high" group in another, depending on the cutpoint chosen.

In a review on P53 as prognostic factor for bladder cancer, Malats et al. found cut-off values ranging between 1% and 75% to define nuclear overexpression.[15] Accordingly, they concluded: "That a decade of research on P53 and bladder cancer has not placed us in a better position to draw conclusions relevant to the clinical management of patients is frustrating".

Thus, forcing cutpoints to fit the data may not only lead to misleading conclusions but may also reduce the usefulness of the results for making clinical decisions. Obviously, in observational studies, several factors can influence the outcome, and a multivariable analysis would be needed. In addition to investigating the functional form for a continuous variable, the researcher must decide which other variables to include in the statistical model. For the breast cancer example, see Sauerbrei and Royston[13] for more detail, and for background and basic issues for interpreting and reporting results from multivariable analyses, refer to Valveny and Gilliver.[16]

Typical weaknesses of statistical analyses

Our example illustrated only one serious problem in statistical analysis. Many other weaknesses have been identified, including:[10]

- inappropriate or inefficient study design
- inappropriate, inefficient, or outdated choice of statistical methods
- misapplication of a valid method
- interpretation problems, including misinterpretation of *P* values, over-confidence in results, misleading interpretation of parameter estimates, bias, and confounding
- reporting problems, including inadequate details for methods and results

Although some methodological errors relate to the failure to grasp some complex or subtle statistical issues, problems in applying even simple methods are widespread (for further details and examples, see Sauerbrei et al.[10] and Lang and Altman[17]).

The need for guidance in planning and analysing observational studies

During the last two decades, several initiatives have been started with the goal of improving research in the health sciences. Transparent and complete reporting is a prerequisite for judging the usefulness of data and interpreting study results in an appropriate context. Reporting guidelines have been developed for many different types of studies. These can be found on the EQUATOR network website (www.equator-network.org/), which serves as a repository of these guidelines and assists in the development of reporting guidelines.[18] The STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) statement provides excellent guidelines for reporting observational studies,[19] and the guiding principles for reporting statistical methods and results were recently published.[17]

Because of the problems in analysing observational studies, guidance on the advantages and disadvantages of competing statistical strategies is needed.[10] For various reasons, this is much more difficult than generating reporting guidelines. In addition, suitable guidance must be tailored to the experience and statistical knowledge of the user, which can vary widely.

The STRATOS initiative

Understanding and overcoming the formidable challenges in designing and analysing observational studies requires a broad-based,

Table 1. Topic groups and their chairs

Topic Groups		Chairs
1	Missing data	James Carpenter (UK), Katherine Lee (Australia)
2	Selection of variables and functional forms in multivariable analysis	Michal Abrahamowicz (Canada), Aris Perperoglou (UK), Willi Sauerbrei (Germany)
3	Initial data analysis	Marianne Huebner (USA), Saskia le Cessie (Netherlands), Werner Vach (Germany)
4	Measurement error and misclassification	Laurence Freedman (Israel), Victor Kipnis (USA)
5	Study design	Suzanne Cadarette (Canada), Mitchell Gail (USA)
6	Evaluating diagnostic tests and prediction models	Gary Collins (UK), Carl Moons (Netherlands), Ewout Steyerberg (Netherlands)
7	Causal inference	Els Goetghebeur (Belgium), Ingeborg Waernbaum (Sweden)
8	Survival analysis	Michal Abrahamowicz (Canada), Per Kragh Andersen (Denmark), Terry Therneau (USA)
9	High-dimensional data	Lisa McShane (USA), Joerg Rahnenfuehrer (Germany)

international group of statistical experts who are also involved with real-world applications. This is the driving vision behind the STRATOS initiative (http://www.stratos-initiative.org), which was launched in 2013 at the 34th conference of the International Society of Clinical Biostatistics (ISCB).[10] STRATOS remains affiliated with the society and had dedicated sessions or mini-symposia at each annual meeting from 2013 to 2016.

STRATOS brings together methodological researchers in several areas of statistics essential for analysing observational studies. These experts have largely complementary knowledge, which allows STRATOS to address complex challenges in the design and analysis of observational studies. STRATOS works to develop, validate, and compare state-of-the-art methods for topics relevant to many kinds of statistical analyses.

Because there is a finite pool of experienced statisticians, many analyses are conducted by researchers with limited statistical literacy and experience. Consequently, the ultimate objective of the STRATOS initiative is to develop guidance for data analysts and researchers with different levels of statistical training, skills, and experience. The initiative considers three levels of statistical knowledge: low (level 1), experienced (level 2), and expert in a specific area (level 3).

Our initial goal is to develop guidance for experienced statisticians (level 2), which involves drafting reviews of methods used in the literature and providing empirical evidence to assess and compare approaches, with the goal of providing arguments for state-of-the-art methodology.

The guidance is informed by a recent list of recommendations for how to improve the uptake of novel methods.[20] It will cover practical issues such as potential pitfalls of inappropriately using "conventional" methods; criteria for choosing appropriate, validated methods that can overcome specific challenges; and software for implementing these advanced methods. The level 2 guidance will then be adapted to researchers with weaker statistical knowledge, which includes most clinicians and medical students (level 1), while experts in specific areas (level 3) will work to identify current gaps in knowledge and improve, validate, and compare existing methods.

STRATOS currently has nine topic groups (TGs) (Table 1), all of which include 8 to 12 members. Further details are available in Sauerbrei et al 2014[10] and on the STRATOS website. Ten cross-cutting panels have been created to coordinate the activities of different TGs, share best research practices, and disseminate research tools and results across TGs (Table 2). These panels address common issues such as creating a glossary of statistical terms, giving advice on how to conduct simulation studies, and setting publication policies for the initiative. The recommendations of the cross-cutting panels are intended to support, integrate, and harmonise work within and across the TGs and to increase transparency in producing guidance. Interested colleagues can apply to become a member of one or two TGs or panels at http://www.stratos-initiative.org.

Summary and outlook

Although substantial progress has been made in designing and analysing data from clinical and epidemiological studies, real-world application lags far behind the advances. This is largely because most researchers have limited knowledge and experience in using advanced statistical methods and software, and even experts can disagree on how best to analyse complex study data, with no consensus on "state-of-the-art" methodology. The STRATOS initiative aims to fill this gap by developing guidance and tools for applying statistical methods for observational research. This is an important step in improving evidence-based decision-making about healthcare.

The STRATOS initiative began in 2013 with about 40 members and, despite a lack of specific funding, has grown to more than 80 members from 16 countries in 2017. Work, research, discussions, and activities are ongoing in nine key relevant areas. Much research, in particular simulation studies, is needed to assess competing statistical approaches. STRATOS's structure is designed to make the resulting guidance broadly useful, but collaboration with clinicians, applied researchers, scientific societies, and related projects and initiatives is needed.

The emergence of "big data" is an additional driver for STRATOS. Big data pose particular challenges and opportunities, and it encompasses diverse areas and data sources. Because of this complexity, STRATOS has decided not to have a big data topic group but instead to encourage all TGs to consider it in their work.

To improve statistical methodology and its transparency, statistical reseachers must put more emphasis on comparing competing strategies and must generate evidence to support state-of-the-art methodologies. They must also provide guidance that is appropriate for the large community of people who analyse and consume data, who have a wide range of statistical knowledge and experience.

If you are interested in the work of the STRATOS initiative or would like to participate, please visit us at http://www.stratos-initiative.org/.

Table 2. Panels, their chairs, and co-chairs

Panels	Chairs and Co-Chairs
Membership	James Carpenter (UK), Willi Sauerbrei (Germany)
Publications	Bianca De Stavola (UK) , Mitchell Gail (USA), Petra Macaskill (Australia), Stephen Walter (Canada)
Website	Joerg Rahnenfuehrer (Germany), Willi Sauerbrei (Germany)
Glossary	Simon Day (UK), Marianne Huebner (USA), Jim Slattery (UK)
Simulation Studies	Michal Abrahamowicz (Canada), Harald Binder (Germany)
Contact Organisations	Douglas Altman (UK), Willi Sauerbrei (Germany)
Literature Review	Gary Collins (UK), Carl Moons (Netherlands)
Data Sets	Hermann Huss (Germany), Saskia Le Cessie (Netherlands), Aris Perperoglou (UK)
Knowledge Translation	Suzanne Cadarette (Canada), Catherine Quantin (France)
Bibliography	To be determined

Conflicts of Interest and Disclaimers

The authors declare no conflicts of interest.

Acknowledgement

The authors thank Dr Phillip Leventhal for help in editing this article.

References

1. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "Optimal" cutpoints in the evaluation of prognostic factors. J Natl Cancer Inst 1994;86:829–35.
2. Ioannidis JP. Why most published research findings are false. PloS Med 2005;2:E124.
3. The Lancet. Research: increasing value, reducing waste [series]. 2014. Available from: www.thelancet.com/series/research.
4. Kleinert S, Horton R. How should medical science change? Lancet 2014; 383:197–198.
5. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009;374:86–9.
6. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 2014;383: 166–175.
7. Sauerbrei W. Prognostic factors – confusion caused by bad quality of design, analysis and reporting of many studies. In Bier H , editor. Current research in head and neck cancer. Advances in oto-rhino-laryngology. Basel, Karger 2005;62:184–200.
8. Trouble at the lab. The Economist 2013 Oct 18.
9. Vickers A. Interpreting data from randomized trials: the Scandinavian prostatectomy study illustrates two common errors. Nat Clin Pract Urol 2005;2:404–5.
10. Sauerbrei W, Abrahamowicz M, Altman DG, le Cessie S and Carpenter J on behalf of the STRATOS initiative. STRengthening Analytical Thinking for Observational Studies: the STRATOS initiative. Stat Med 2014;33:5413–32.
11. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 2006;25:127–41.
12. Van Walraven C, Hart RG. Leave 'em alone – why continuous variables should be analyzed as such. Neuroepidemiology 2008;30:138–9.
13. Sauerbrei W, Royston P. Continuous Variables: To categorize or to model? In: Reading C, editor. The 8th International Conference on Teaching Statistics – Data and Context in statistics education: Towards an evidence based society. International Statistical Institute, Voorburg. 2010.
14. Royston P, Sauerbrei W. Multivariable model-building – a pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Wiley, Chichester 2008.
15. Malats N, Bustos A, Nascimento CM, Fernandez F, Rivas M, Puente D et al. P53 as a prognostic marker for bladder cancer: a meta-analysis and review. Lancet Oncol 2005;6:678–86.
16. Valveny N, Gilliver S. How to interpret and report the results from multivariable analyses. Med Writ 2016;25:37–42.
17. Lang TA, Altman DG. Statistical analyses and methods in the published literature: The SAMPL guidelines. Med Writ 2016; 25:31–5.
18. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. BMC Med 2010;8:24.
19. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP for the STROBE initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Epidemiology 2007;18:800–4.
20. Cadarette SM, Ban JK, Consiglio GP et al. Diffusion of innovations model helps interpret the comparative uptake of two methodological innovations: co-authorship network analysis and recommendations for the integration of novel methods in practice. J Clin Epidemiol 2017; 84:150–60.

Author information

Dr Willi Sauerbrei is a Senior Statistician and Professor in Medical Biometry at the University of Freiburg. His main interest is in various issues of model building. With Patrick Royston, he developed the multivariable fractional polynomial approach and extensions of it. He initiated and is the Chair of the STRATOS initiative.

Dr Gary S. Collins is a Professor of Medical Statistics at the University of Oxford, he is also the Deputy Director of the UK EQUATOR Centre. He is a co-founding Editor of the BMC journal Diagnostic and Prognostic Research and is a Statistics Editor for the *BMJ*.

Dr Marianne Huebner is an Associate Professor at Michigan State University, MI, USA. Previously she worked at Mayo Clinic, MN. Her research is in biomedical statistics and statistical modeling for health care outcomes. She is a Statistics Editor for the *Journal of Cardiovascular and Thoracic Surgery.*

Stephen Walter is a medical statistician at McMaster University, where he is an Emeritus Professor. Dr Walter works collaboratively on methods and applications dealing with: design and analysis of medical research studies; risk assessment and communication; diagnostic and screening data; and spatio-temporal variation in health.

Dr Suzanne M. Cadarette is Associate Professor of Pharmacy, University of Toronto; Senior Adjunct Scientist, Institute for Clinical Evaluative Sciences; and Adjunct Associate Professor of Pharmacy, University of North Carolina. She is co-chair of the STRATOS KT Panel and of TG5 (Study Design).

Michal Abrahamowicz is a James McGill Professor of Biostatistics at McGill University, in Montreal, Canada. His research involves development and validation of new, flexible models for survival analysis, and for controlling for different biases in observational studies, with applications in pharmaco-epidemiology. He is a co-chair of the international STRATOS initiative.