

From Willi Sauerbrei, Michal Abrahamowicz and Saskia Le Cessie, for the STRATOS initiative

According to the general paradigm of modern sciences, statistical methods are key to translate raw empirical data into new insights in, and deeper understanding of, complex processes affecting e.g. human health, the economy, or environment. Yet, the complexity of such processes, and of the observable data they generate, create numerous analytical challenges. In the 21<sup>st</sup> century, parallel progress in the theory of mathematical statistics and computational resources and technology led to dynamic developments in statistical methodology, resulting in a large number of increasingly complex, ever more flexible, statistical techniques and models that address several complexities frequently encountered in analyzing real-life data. Unfortunately, many of these important developments are ignored in every-day practice of data analysis. Consequently, the design and analysis of recent, often complex and costly, observational studies published in high-impact medical or social sciences journals often exhibit serious weaknesses; frequently resulting in misleading inferences and incorrect conclusions.

Formulating and overcoming these formidable challenges requires a well structured, highly interactive collaboration between a large, international group of statistical experts, whose research combines development of new methodology with collaborative research on real-life applications and whose joint, complementary expertise covers different sub-areas of statistical research. The need for such collaboration arises due to a combination of (i) ever increasing complexity and variety of analytical challenges encountered in observational studies, together with (ii) the increasing trend for narrow specialization, necessary to achieve cutting-edge novel developments in modern statistics. At present, particular challenges are being addressed by leading authorities in different areas of statistical research, but little effort is invested in combining the results of these separate developments and ensuring their material impact on the practice of data analysis. This situation provided the motivation for, and driving vision behind, the STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative.

The ultimate objective of the STRATOS initiative is to develop guidance documents for data analysts and researchers with different levels of statistical training, skills and experience. Specifically, we have identified three levels of statistical knowledge. Initially, we are working to develop guidance documents for experienced statisticians (level 2), which involves drafting reviews of state-of-the-art methods relevant for a number of specific topics (listed below in the table). The guidance documents will cover such practical issues as potential pitfalls due to inappropriate use of 'conventional' methods, the criteria for choice of appropriate, validated methods able to overcome specific challenges, and software to implement these advanced methods. The level 2 guidance results will then be simplified for analysts with low statistical knowledge, including e.g. clinicians and medical students (level 1). In parallel, experts in a

specific area (level 3) will help to identify current gaps in knowledge and try to improve, validate and/or compare the existing methods. For more details see Sauerbrei et al (2014) and the STRATOS website <http://www.stratos-initiative.org/>.

### **STRATOS as an intellectual child of ISCB**

The need to develop guidance for the analysis of observational studies was the first item of the Epidemiology SC (EpidSC) meeting at ISCB2011 in Ottawa. None of the ten present members knew any suitable document and Willi Sauerbrei (WS) proposed that the SC should pursue this challenge. His suggestion was based on very positive experiences in developing reporting guidelines to enhance quality and transparency of health research. During the last two decades reporting guidelines were developed for many types of studies, with the EQUATOR network (<http://www.equator-network.org>) acting as an umbrella. EpidSC mandated WS to start a project aiming at guidance documents for the analyses of observational data. Several colleagues (most being ISCB members) joined a steering group and the STRATOS initiative was launched with a very successful mini-symposium at the ISCB 2013 meeting in Munich. At each of the ISCB meetings in 2014–16 STRATOS had dedicated invited sessions or mini-symposia (website provides details and slides from all talks). When launching the initiative in August 2013, STRATOS had 45 members. Despite being without official core funding (except use of research funds available to individual members and ISCB support for invited speakers), by November 2016 membership has already increased to more than 80 researchers from 16 countries on 4 continents. This impressive increase reflects the interest of methodologists to work on guidance documents for the analysis of observational studies and the large audiences at all STRATOS sessions and symposia reflect the biostatistical research community's need for such documents.

In July 2016, the STRATOS initiative had its first General Meeting at the Banff International Research Station (BIRS) in the Canadian Rocky Mountains, co-organized by Willi Sauerbrei and Michal Abrahamowicz, and supported by the BIRS grant. During the week-long BIRS workshop entitled 'Developing a Comprehensive, Integrated Framework for Advanced Statistical Analyses of Observational Studies' 38 STRATOS members and 7 research trainees discussed topic-specific and general STRATOS-wide projects (website provides a detailed report). This important meeting informed and stimulated further activities in the following weeks.

Invited, video recorded, STRATOS session at the International Biometric Conference (<https://biometricconference.org/>), STRATOS Mini-symposium at ISCB2016 (<http://www.iscb.info/ISCB2016.html>), special STRATOS session at the HEC2016 conference (Health—Exploring Complexity: An Interdisciplinary Systems Approach; <http://www.hec2016.eu/>).

...

Although STRATOS collaborates with several societies and projects, it is a brainchild of ISCB and has closest connections with ISCB and its EpidSC. Many of the ISCB members work with observational data and their numbers are expected to grow as more and more health care data and data from large registries are available for research. The great interest of ISCB members for observational studies and the STRATOS initiative can be seen from the large attendance of the STRATOS sessions at ISCB meetings. Typical ISCB members may be considered experts (level 3) in one or two topics, on which their research focuses, and level 2 analysts in other topics. The majority of the EpidSC members are also STRATOS members and the majority of the STRATOS steering group members have been active in ISCB, as President, ExCom members, members of SCs and SPCs, presenters of courses and invited speakers.

We expect the ISCB Educ SC (chaired until recently by Catherine Quantin, a STRATOS member) will play an important role in knowledge translation of STRATOS papers and guidance documents, planned for the next year. The STRATOS knowledge translation panel is enthusiastic to help promote activities and encourage adoption of guidance documents among other societies. The overlap in members of EpidSC and STRATOS will ensure that STRATOS documents will be presented at ISCB meetings and educational workshops supported by ISCB; an example is the presentation by WS to the Romanian National Group (see the RNG report 2015, ISCB News # 60).

#### **STRATOS topic groups and cross-cutting panel**

STRATOS currently has 9 topic groups (TGs). Topic groups and their chairs are listed in the table. All groups have about 8 to 12 members.

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

To co-ordinate the activities of different TGs, to share best research practices, and disseminate research tools and results across TGs, several cross-cutting panels have been created recently. These panels aim to address 'generic' issues, such as simulation studies, and publications policies, and to develop recommendations and coordinating the efforts of the individual TGs. Recommendations aim to support, integrate and harmonize work within and across the TGs. They will also help increase transparency in deriving final guidance documents for the entire STRATOS initiative.

The website lists current membership, provides short summaries of topic groups and panels, and includes the link to the detailed report of the BIRS meeting in July 2016.

#### **Recent activities, future challenges and plans**

Most STRATOS topic groups will circulate draft guidance documents, focusing on the issues fundamental for their respective areas, by the end of 2018. Some TGs have already derived roadmaps for specific steps and projects necessary to achieve this goal, including validation and comparison of alternative methods in comprehensive, contextually relevant simulation studies.

The emergence of 'Big Data' is a key driver for STRATOS, as this poses particular challenges and opportunities across the spectrum of statistical research and applications. However, the term 'Big Data' encompasses as diverse areas and data sources as electronic health records, large administrative and insurance-oriented databases, related registry data, -omics, and imaging data. Reflecting this diversity, STRATOS deliberately decided not to have a 'Big Data' topic group, but instead to encourage all TGs to consider how their work relates to, can be motivated by, and/or adapted to, specific challenges induced by the type of 'Big Data' relevant to their focus. This process has started in many TGs. The High Dimensional Data TG9 is concerned with the specific analytical problems that arise with -omics data, where the number of variables is typically far greater than the number of study subjects.

We plan to organize a second general STRATOS meeting in 2018, where a key focus will be on novel research projects addressing analytical challenges frequently encountered in real-life observational studies, including the emerging challenges induced by Big Data.

The STRATOS initiative will continue to inform all ISCB members on new insights and guidance documents regarding observational studies.

*Sauerbrei W., Abrahamowicz M., Altman D.G., le Cessie S. and Carpenter J. on behalf of the STRATOS initiative. (2014) STRENGTHENING Analytical Thinking for Observational Studies: the STRATOS initiative. Statistics in Medicine, 33: 5413-5432.*

<http://www.stratos-initiative.org/>