Developing a Comprehensive, Integrated Framework for Advanced Statistical Analyses of Observational Studies

Michal Abrahamowicz (McGill University, Montréal, Canada),
James Carpenter (London School of Hygiene & Tropical Medicine, London, UK),
Willi Sauerbrei (Medical Center, University of Freiburg, Freiburg, Germany)

July 3–8, 2016

1   The STRATOS initiative — motivation, mission, structure and main aims

The overall objective of the 16w5091 Workshop was to jump-start the research activities of the international initiative for STRengthening Analytical Thinking for Observational Studies (STRATOS). This was achieved by providing the first opportunity for a large group of 38 STRATOS members, and seven research trainees, from 12 countries on 4 continents, to meet in-person, exchange ideas, continue with article writing and together develop an operational plan for further development and internal homogenization of the scientific and knowledge translation endeavors of the nine topic groups and several cross-cutting panels. As the attached Workship Scientific Program shows, the majority of sessions were organized following the structure of STRATOS Topic Groups (TGs) or Panels. Therefore, to facilitate understanding of the rather unique character and goals of the workshop, we begin this Report with a short overview of the STRATOS initiative, including its *raison d'être*, overall mission, internal structure and main objectives.

Lack of statistical expertise is now recognized as a significant brake on scientific progress across a wide range of empirical sciences. For instance, a recent article in The Economist (19/10/2013), 'Unreliable research: Trouble at the lab.' commented: "*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.*"

According to the general paradigm of modern sciences, statistical analysis methods are key to translate raw empirical data into new insights in, and deeper understanding of, complex processes affecting human health, the economy, environment, and many other phenomena studied in different branches of science. Yet, the complexity of such processes, and of the observable data they generate, create numerous analytical challenges. In the 21st 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 allow researchers to account for several complexities frequently encountered in analyzing real-life data. Unfortunately, many of these important developments are ignored in every-day practice of data analysis, including analyses reported in influential publications in high-impact medical or social sciences journals. Consequently, the design and analysis of recent, often complex and costly, observational studies of human health and welfare often exhibit

1

serious weaknesses. This leads to misleading inferences, which may, in turn, adversely affect the effectiveness or safety of different treatments, social or economic policy programs etc.

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. Indeed, such need arises due to a combination of (i) ever increasing complexity and variety of analytical challenges encountered in the majority of important observational studies, together with (ii) the increasing trend for narrow specialization, necessary to achieve cutting-edge novel developments in modern statistics. Together, these two trends imply that no single university-based (bio-)statistics department and no team of (bio-)statisticians (working in even the largest private or public research institutes or, e.g., pharmaceutical companies) is able to ensure state-of-the-art expertise regarding even a reasonable fraction of analytical challenges encountered in real-life applications. 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 STRATOS initiative was launched in 2013 at the 34th conference of the International Society of Clinical Biostatistics (ISCB). It is connected to this society and had dedicated invited sessions or mini-symposia at all ISCB annual meetings in 2013–16. The initiative brings together leading, internationally recognized, methodological experts in several areas of statistics essential for the analyses of observational studies, who—at present—are grouped in nine Topic Groups (TGs), each focusing on a different, highly relevant area of statistical research. The experts' joint, largely complementary, knowledge allows the initiative to address complex analytical challenges in the design and analysis of observational studies, by developing, validating and comparing state-of-the-art methods for various topics. To increase the impact of our endeavours on empirical research, individual topic groups are working to summarize our findings by developing practical guidance regarding 'best practice' to address a particular set of analytical challenges (e.g. handling of missing data or dealing with right-censored time-to-event data). The guidance covers such practical issues as e.g. the awareness of potential pitfalls due to inappropriate use of 'conventional' methods, the choice of appropriate, validated analytical methods able to overcome specific challenges, and software that can be used to implement these advanced methods.

We are entering the era of 'big data' with automated collection of very large amounts of data and the paradigm of empirical sciences shifting toward 'data science'. In a recent Editorial, David Hand (*J Roy Statist Soc (JRSS)* A (2016), 179: 629–631) stresses the importance of distinguishing between two types of activity relating to big data. The first involves primarily data manipulation: sorting, searching, matching, and so on. Examples include online route finders and apps for updated status of bus and train traffic, with the associated issues addressed mostly by computer scientists and mathematicians. The second type of big data activity seeks to go beyond the data at hand, with the ultimate goals being either prediction of future data, or understanding of the mechanisms and processes that have generated the collected data. Achieving these goals will rely primarily on state-of-the-art statistical and machine learning methods.

However, 'big data' will *not* help answering such essential prognostic or etiology questions if researchers use designs and statistical methods which are unsuitable, e.g. by being unable to account for the complexity of the underlying dynamic multi-factorial processes. Therefore, it is of central importance to gain knowledge about advantages and disadvantages of alternative statistical approaches, and their dependence on the data structure. Equally important is to develop, validate and explain to end-users state-of-the-art methods that can address frequent limitations (e.g. missing data, measurement errors, unmeasured confounding) and complexities of the data (e.g. non-linear relationships of continuous variable with the outcome, time-varying effects, mediation). These challenges are relevant for small, intermediate and big data sets, and the STRATOS initiative aims to address all of them. In the health sciences, studies with a relatively small sample size and a very large number of potential (genetic) predictors ($p >> n$) have become popular. Obviously, the analysis of such studies raises many methodological issues. To cope with such problems we have recently created 'High dimensional data' TG9. The work of this group is summarized in Section 3.9, and other interactions at Banff relating to 'Big Data' are described in Section 6.1.

One of the fundamental objectives of the STRATOS approach is to develop guidance documents for data analysts and researchers with different levels of statistical training, skills and experience. Specifically, we have indentified three levels of statistical knowledge, each of which would require a somewhat different

	STRATOS 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

Table 1: Topic groups and their chairs

type of targeted guidance document, and we have outlined the main criteria to be used when developing the guidance documents aimed at the analysts at each level. Initially, we are working to derive guidance documents for experienced statisticians ('level 2'), which requires work on state-of-the-art documents for each specific topic group. For each topic considered (see next section) several analytical strategies have been proposed in the statistical literature, but knowledge about their properties and relative strengths and weaknesses is often insufficient, as meaningful comparisons are rare and evidence based guidance documents are lacking. For more details see Sauerbrei *et al*, (*Statistics in Medicine*, 2014, **33**, 5413–5432) and the STRATOS website http://www.stratos-initiative.org/

Several members of STRATOS have recently worked on various guidelines for reporting in the health sciences, research now coordinated by the equator network (http://www.equator-network.org/). Positive experiences in reporting guideline initiatives have been an important driving force and a source of inspiration for the discussions undertaken during the BIRS workshop.

1.1 STRATOS topic groups

August 2013, when the STRATOS initiative was launched, it had seven Topic Groups (TGs) and 45 members. By July 2016, these numbers have increased to 9 TGs (see Table 1), with a total of 82 members, from 15 countries.

1.2 Cross-cutting panels

To co-ordinate the activities of different topic groups (TGs) , share best research practices, and disseminate research tools and results across TGs, several cross-cutting panels have been created in 2015–16. These panels aim to address the 'generic' issues, common to all or most of the TGs, by developing recommendations (sometimes rather loose as for simulation studies, sometimes more strict as for STRATOS publications) 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.

2 The BIRS Workshop

2.1 Participants.

Between the official launching of the STRATOS initiative in August 2013 and the Spring of 2016, some of the members met during the annual ISCB conferences, and each of the topic groups (TGs), as well as the steering group conducted more or less regular teleconferences, in addition to (separate) small face-to-face, 1–2 days long, meetings of subgroups of 3–6 members of some of the TGs. However, the BIRS workshop offered the first opportunity for the members of all 9 TGs and the panels to meet together, and discuss both TG-specific and general STRATOS-wide projects over the extended period of five working days, often followed by less formal evening meetings of subgroups interested in particular methodology, or organizational issues, related e.g. to strategic planning or STRATOS *modus operandi*. Together these varied, highly interactive activities were essential to start the complex, long-term process whose ultimate objective is to develop a comprehensive, integrated framework for the statistical analysis of observational studies.

The popularity, and interest, of the international statistical research community in the workshop were reflected in both (1) the fact that we have filled all the 42 available places at BIRS, and in addition had three non-resident observers; and (2) the meeting's truly global character. Indeed, the 45 participants/observers came from 12 countries on 4 continents (most participants came from Canada, USA and Europe but Asia (J. Freedman from Israel) and Australia (K. Lee) were also represented). The participants represented the full spectrum of career stages, including (i) leading international authorities in different fields of statistics and biostatistics; (ii) several mid-career and junior faculty members; as well as 7 research trainees (2 post-doctoral fellows and 5 PhD students). About 40% of participants were female.

2.2 Overview of the presentations and discussions

On Monday morning, we started with a general introduction to the initiative, overview of its long-term objectives and mode of operation, and the main aims of the BIRS meeting. This was followed by 4 keynote talks by internationally recognized authorities in selected areas covered by the STRATOS initiative: Per Kragh Andersen (survival analysis, TG8), James Carpenter (missing data, TG1), Frank Harrell (selection of variables and functional forms in multivariable analysis, TG2), and Els Goetghebeur (causal inference, TG7). To jump-start the discussions over the next 4 days, the four keynote speakers provided a broad overview of the state of research in their respective areas and identified some outstanding methodological challenges, as well as pointed out some of the most frequent mis-conceptions and errors in the current practice of data analysis.

The key component of the meeting involved presentations and highly interactive discussions about the activities of nine TGs. TG-specific presentations, usually given by 2 or 3 members of a given TG focused on: (i) overview of methods currently used in everyday practice of data analysis, and their potential limitations, and/or (if applicable) of the existing guidelines, tutorials or methodological review papers covering the methods relevant to a particular TG, (ii) main analytical issues that should be clarified in future guidance documents, (iii) progress already achieved by a given TG, (iv) outstanding challenges, some of which may require either primary research on developing new methods or secondary simulation-based research on validation and comparison of alternative existing methods, (v) short- and mid-term plans, including outlines for future papers, as well as (vi) identification of future collaborative links to some of the other STRATOS TGs, in order to tackle more complex issues that will require combining cutting-edge methods developed by individual TGs (e.g. handling of missing data on unmeasured confounding in survival analysis (TG1, TG7, TG8), or selection of variables and functional forms in high-dimension analyses (TG2 and TG9). All TGs worked on a roadmap for guidance documents, which included intended publications for the next 12 months. Key members were identified for most scientific projects. More details on the presentations given, and conclusions reached by members of individual TGs, are found in Sections 3.1–3.9 below.

In addition, the STRATOS Steering Group has more recently decided to create various cross-cutting 'Panels', each of which was given a mandate to cover specific strategic issues that are pertinent to many TGs, and/or to the operation of the STRATOS initiative as a whole. Given the cross-cutting character of the panels, the BIRS workshop offered an unique opportunity for direct in-person discussions of the panels missions and their *modus operandi*. Therefore, the panels were given an important role at the meeting. Accordingly, during the 5 days of the BIRS meeting, the necessity and role of each of the 10 panels were discussed, usually after

a detailed presentation from one of the chairs. General discussions were often followed by in-person meeting of members of different TGs interested in a given panel, and discussions in smaller groups. Several panels recruited new members (Sections 4.1–4.7 of the Report provide a summary of the activities of the respective panels during the BIRS workshop).

To allow more detailed discussions and time for outlining the content of future manuscripts within TGs, as well as discussions between members of different TGs and/or panels, we also decided dedicate a substantial amount of time to separate meetings in smaller groups. Often discussions were continued in the evenings. Results of small groups discussions were presented and discussed on Friday morning, in a general session that summarized the conclusions of the BIRS meeting and provided an outlook for main activities planned for the next two years.

2.3   Summary of the BIRS workshop achievements and plans for future

There was a general consensus that the BIRS workshop was very successful: (i) each of the nine topic groups made substantial progress in shaping and concretizing their research plans, and was able to outline or even draft the manuscript(s) (more than 15 TG-specific manuscripts are currently in progress, at various stages of completion, across the STRATOS initiative), (ii) first multi-disciplinary projects, tackling complex analytical challenges and issues relevant for more than one TG were outlined, (iii) members of the initiative refined the purpose of, and agreed members for, the cross-cutting panels, and (iv) progress was made on an effective dissemination strategy including an enhanced, interactive, website and social media presence, as well as concrete steps to build long-term partnerships with various societies and organization including those that involve mostly (a) statisticians or (b) end-users of statistical methods.

All participants expressed the strong wish having more regular meetings in smaller groups (TG meetings, joint meetings of members from two or more TGs and/or one or two panels with somewhat overlapping mandates (e.g. Knowledge Translation vs Publication panels). There was also an overwhelming consensus that, given the achievements of the BIRS meeting, it will be essential to organize the next general meeting of the STRATOS team in about two years. Several colleagues suggested that an ideal location for the next meeting, in 2018, will be Oberwolfach (Germany), whose research station has a mission similar to BIRS. Accordingly, four of the Steering Group members started working on the new application in the days after the BIRS meeting and, on July 31, 2016 have submitted an application for a workshop in 2018 in Oberwolfach, entitled 'Next stages in Developing an Integrated Framework for Advanced Statistical Analyses of Observational Studies'.

3   Summary of Progress and Plans by 9 Topic Groups

All topic groups benefitted immensely from face-to-face meetings, with many STRATOS members meeting for the first time. This was vital for both cementing interpersonal relationships and establishing new long-term research collaborations, the continuation of which will be almost exclusively by email and teleconferences.

Project management and plans for the first topic group-specific STRATOS manuscripts, were discussed within each TG; it was agreed that authorship will broadly follow the ICMJE guidelines. In general, all TG members who expressed an interest are invited to actively contribute to a given project, and input from other TG members should be sought at various stages, e.g. before presentations and for revisions of manuscript drafts. More specific details of the protocol for STRATOS publications are given below, in a sub-section 4.1 on the Publication Panel.

TG-led papers were finalized, revised and developed, as summarized below:

3.1   TG1: Missing data

Discussions and interactions

Five members of the group were present (including the co-chairs); it was the first time some of us had met, and good collaborative relations were established. We reviewed progress so far, and decided to re-focus our efforts as outlined below.

Research: current

Our first paper, led by Rod Little, will be aimed at a 'level 1' audience, and will detail when analysis of records with no missing data alone is likely to be appropriate. Preliminary discussions with the New England Journal of Medicine suggested they would be open to considering a short 'research methods' article on this topic.

The second paper, led by Kate Lee, with aspects relevant for both level 1 and 2 researchers will outline the proposed framework for the conduct and reporting of the analysis of observational studies with missing data. This framework was presented by James Carpenter in the plenary session, and refined following the ensuing stimulating discussions.

These papers should be drafted by January 2017.

Research: future

Beyond this, we also planned a paper led by James Carpenter aimed at 'level 2' researchers, presenting and comparing a wide range of the different approaches to analyzing data that are missing at random, using an extract from publicly available UK youth cohort study.

3.2 TG2: Selection of variables and functional forms in multivariate analysis

Discussions and interactions

The plenary talk by Frank Harrell stimulated an ongoing discussion about the utility of variable selection *vis-a-vis* different approaches to variables selection.

Research: current

The meeting allowed the plans of the group to be consolidated as follows:

- Michal Abrahamowicz and Gary Collins (TG6) will lead on a literature review of the relevant methods currently used in applied research, with a draft circulated to the group by December 2016;

- Aris Perperoglou and Matthias Schmid will lead the review and assessment of the splines packages available in R, circulating a draft to the group by November 2016;

- Pre-modelling data analysis was identified as an important, and somewhat neglected, area and Matthias Schmid and Werner Vach will co-ordinate a draft of a level 1 paper to be circulated in December 2016.

Research: future

- The group agreed that the conflicting, and often missing, information about the pros and cons of the many spline approaches, in the current literature, is confusing to researchers and calls for a systematic assessment and clarification. Aris Perperoglou, Matthias Schmid and Willi Sauerbrei will draft a level 2–3 article for the group by mid 2017.

- Spikes at zero are a challenge for many methods, and Heiko Becher and Frank Harrell will draft a level 2 manuscript evaluating the suitability of methods in this setting for Summer 2017.

- Willi Sauerbrei, Matthias Schmid, Doug Altman and Aris Perperoglou will work on a level 1 manuscript assessing how analysts handle modeling of the functional forms for continuous prognostic factors (whether they categorize them, etc.)

3.3 TG3: Initial data analysis

Discussions and interactions

Three members of the TG, Heike Hoffmann, Marianne Huebner, Werner Vach, were present. The group agreed to drop 'Descriptive' from the name, and focus on initial data analysis of observational health care

studies. Future aims for TG3 are data visualization, skewed covariates, literature review, IDA reporting guidance, IDA for longitudinal data, and electronic health records.

The group took full advantage of the opportunity for face-to-face discussions with other topic groups, in particular with:

- TG9 (High dimensional data analysis), to exploit approaches of common interest. As a result, Lara Lusa (TG9) has joined TG3;

- TG1 (Missing data), to discuss the visualization of missing data patterns, and how this can be used to show predictors of data being missing.

- TG2 (Variable selection), leading to planning a paper 'What you should think about before starting a regression analysis', led by Werner Vach and Matthias Schmid.

Research: current

The following research presentations were given by group members at Banff:

- Marianne Huebner: 'Overview of framework and reporting of initial data analysis';

- Heike Hoffmann: 'Data visualization for initial data analysis'

These informed discussions leading to finalization of a level 3 manuscript, entitled 'A contemporary framework of initial data analysis', which has been submitted;

As mentioned above, plans were developed for a paper 'What you should think about before starting a regression analysis', led by Werner Vach and Matthias Schmid.

A paper on data visualization for initial data analysis led by Heike Hoffmann and Dianne Cook is in progress.

Research: future

The following projects are being developed:

- a literature review

- initial data analysis for longitudinal data, especially from electronic health records

- issues with skewed data

3.4 TG4: Measurement error and misclassification

Discussions and interactions

Five members were present in Banff: Veronika Deffner, Larry Freedman, Ruth Keogh, Victor Kipnis, and Pamela Shaw. It was agreed to:

- nominate TG members for the panels as follows: Simulation Panel—Victor Kipnis (secondary representative Pamela Shaw); Website—Ruth Keogh; Publications—Pamela Shaw;

- use, and advocate to the Datasets Panel, the NHANES, OPEN, IDATA and Whitehall datasets, and

- Suggest to the Organizations Committee linking with the ISEE, Farr Institute, ICDAM, and Epi-Congress (SER).

Research: current

Previous plans to write three papers by Spring 2017 were consolidated as follows:

- a level 1–2 report of our literature survey on the use of methods to adjust for measurement errors in four areas of epidemiology;

- a level 2 guidance paper on measurement error and misclassification of variables in epidemiology, and

- a guidance paper on dietary measurement errors for nutritional epidemiologists.

Research: future

- Discussions with Mitchell Gail (TG5, Design) elaborated several areas where improved design could address measurement error issues. The two groups plan to collaborate on this in 2017.

- Discussions with Lisa McShane (TG9, High-dimensional data), and James Carpenter (TG1, Missing data) suggested fruitful areas of collaboration, and these will be developed further in 2017.

3.5   TG5: Study design

Discussions and interactions

Suzanne Cadarette joined TG5, and discussions on TG5 representatives on various STRATOS panels were initiated. Working groups were established with the following topic groups where improved designs need to be developed and adopted:

- TG4 (Measurement error): led by Laurence Freedman and Mitch Gail, and

- TG9 (High dimensional data): led by Lisa McShane and Mitch Gail, focusing on use of stored bio-specimens for biomarker studies for prognosis or treatment prediction.

Research: current

The TG has an advanced draft of an overview of design issues in observational studies; it was agreed to strengthen this with three examples:

- Australian weighted cohort with nested case-control design (Elizabeth Williamson);

- GWAS studies (Mitch Gail), and

- Administrative database example to illustrate confounding by indication (Elizabeth Williamson).

The manuscript should be submitted by the end of 2016.

Research: future

The following projects are in the planning stage:

- paper on design aspects of routinely collected data (Elizabeth Williamson, Suzanne Cadarette, and Andrea Rotnitsky)

- design for the phases of (i) development and (ii) validation of risk prediction models in new prognostic factor(s) (Gary Collins, Peggy Sekula).

3.6   TG6 Evaluating diagnostic tests and prediction models

Discussions and interactions

During the meeting, a Letter to the Editor highlighting methodological flaws of a recently published prediction model in the journal *Gynecological Oncology Reports* was written. The letter has subsequently been published (on behalf of TG6):

Van Calster B, Steyerberg EW, Bourne T, Timmerman D, Collins GS, on behalf of TG6 of the STRATOS Initiative. Flawed external validation study of the ADNEX model to diagnose ovarian cancer. *Gynecological Oncology Reports* 2016; DOI: 10.1016/j.gore.2016.09.003

Research: current

- Design for the phases of development and validation of a prognostic model, with TG5 (Design), led by Gary Collins and Peggy Sekula.

Research: future

- Perform a number of systematic reviews evaluating current practice in the medical literature to identify areas where guidance is most needed.

- Write an overview/review paper describing the various approaches for assessing model calibration; provide some guidance for non-statisticians.

- An article on how to compare the performance of alternative prediction models in an external validation.

3.7   TG7: Causal inference

Discussions and interactions

Causal inference is clearly related to a number of the TGs, and this lead to a number of fruitful interactions, in particular with TG1 (missing data), TG4 (Measurement error), TG8 (Survival analysis), and TG9 (High dimensional data)

Research: current

The aim is to provide insight and guidance in this very complex landscape by identifying and publishing important studies, where causal effects have been correctly (or incorrectly) handled:

- a case study on the effects of breastfeeding on the baby's weight was identified,

- discussions with TG8 (Survival) focused on the challenges of phrasing the causal question in the competing risks setting.

Research: future

- Use the breast feeding example to construct a realistic simulation framework.

- Develop the sketch paper on missing data issues in causal inference in collaboration with TG1.

3.8   TG8: Survival Analysis.

Discussions and interactions

All but two TG8 members were present. Members took full advantage of the opportunity for fruitful discussions, laying the foundation for progress that would be been substantially more difficult and time consuming without the BIRS meeting.

Research: current

Discussions focused on the content and structure of the two first TG8 papers, both aimed at the 'level 2' readership (applied statisticians who are not experts in survival analysis). A writing plan was agreed, which should see these two articles submitted to *Statistics in Medicine* in the first half of 2017.

The first paper will concentrate on the relatively common setting where individual members of a (retrospective or prospective) cohort are followed until they are right-censored or developed a single end-point (e.g. 'all-cause mortality'). Following a brief introduction to survival analysis, this paper will provide a detailed discussion of three specific issues, about which there has been considerable research, which has yet to fully permeate practice. These are: (i) modeling of time-varying covariates (i.e. variables which change their value(s) over time); (ii) accounting for time-dependent effects (e.g. non-proportional hazards), and (iii) 'immortal time bias' and related biases.

The second paper will deal with the situation where several end-points are analyzed simultaneously, in a multi-state model, that involves transitions between different health states. These states may be mutually exclusive and/or may follow each other. In the former case a competing risks approach will be described and

promoted. Examples and illustrations will build on Therneau's vignette to his 'survival' package for the R software.

Research: future

Papers for 'level 1' (non statisticians) will be planned at a later stage and will build on the two above-mentioned 'level 2 papers'. These papers for level 1 must be carefully calibrated with the excellent series of four papers from *British Journal of Cancer*, published in 2003 by Altman and others, who provide a broad, reader-friendly overview of (i) many analytical challenges and (ii) most popular statistical methods that may be used to address most of these challenges.

3.9 TG9: High dimensional data

Discussions and interactions

Three members of this, the most recently created, TG were present. Discussions and interactions were very important. Discussions focused on:

- what the notion '*high-dimensional*' should encompass, deciding that as a working definition we would use the term to describe the situation where a large number of variables is available for the analyses.

- deciding that data from molecular medicine (i.e. "-omics" data, like genomics, transcriptomics, proteomics, metabolomics), as well as data from other sources such as electronic health records and comparative effectiveness studies, would be included within the scope of TG9.

- agreeing that the following sub-topics would be used to structure the group's work: Data Pre-processing, Exploratory Data Analysis, Data Reduction, Multiple Testing, Prediction Modeling/Algorithms, Comparative Effectiveness and Causal Inference, Design Considerations, Data Simulation Methods, Resources for Publicly Available High-dimensional Data Sets.

Research: current

An initial project will collate and review issues arising from the analyses of gene expression data from cancer studies that TG members have been involved with and/or have access to, leading to identification of typical, widely used methodologies, their strengths and limitations, and hence guidance for practitioners.

Research: future

Tentative plans include preprocessing methods and data analysis pipelines for specific types of omics data TG members are working with.

4 Cross-cutting Research Panels

Underpinning the initiative are a number of panels, formed from members of the various Topic Groups and some additional STRATOS members whose main contributions to the initiative involve work on specific panels. Panels provide a common scientific framework for the initiative's activities, determine some aspects of its *modus operandi* and provide guidelines regarding issues relevant to all TGs.

As this was the first meeting of the initiative, a part of the Scientific Program was set aside to consolidate the mandates and membership of these panels, which have a key role in the success of the initiative.

Below we present brief reports of the panels' activities during the BIRS workshop.

4.1 Publication panel (PP)

Before the meeting, the PP panel prepared and circulated a draft of the publication policies and procedures for the initiative. This covers issues relevant for peer-reviewed papers, letters, conference presentations and web materials. The panel presented the draft policy in a plenary session, with the focus on rules for authorship of STRATOS publications, as well as responsibilities and timelines of the PP. Following a constructive discussion, the consensus was reached regarding the operational definition of a 'STRATOS publication', and the corresponding standards for internal STRATOS review process.

The full version will be published on the website and initiative members alerted by email. Key requirements agreed upon are:

- Papers should be open access, if at all possible;

- Results need to be reproducible, with both (i) the code for implementation of the analyses, or simulation studies; and (ii) the data used in the analyses, being available to research community at large.

4.2 Glossary Panel (GP)

This panel aims to standardize language, terminology and notation used within the STRATOS initiative's publications and guidance documents. To this end, the GP supports all members helping them to use consistent, clear and unambiguous language in all outputs. In turn, it is hoped this homogenization will positively impact the broader statistical research community.

Prior to the meeting, the panel had prepared a 300 page document of about 800 terms, drawing closely on Simon Day's *Dictionary for Clinical Trials* (2nd edition, Chichester: John Wiley and Sons, 2007. ISBN 047031916X, 9780470319161). The GP will use this as a starting document for the STRATOS glossary. The current version was presented at a plenary session, and suggestions for (i) the structure and searchability, (ii) specific terms for inclusion, and (iii) cross-referencing, were discussed. Members were invited to provide extensions and improvements. The panel will develop plans for incorporating the glossary into the STRATOS website, alongside with user-friendly tools that will allow some highly qualified researchers, to add new terms.

4.3 Review Panel (RP)

This panel is responsible for establishing a common framework for conducting systematic reviews, of both (i) methodological and (ii) applied literature, across all STRATOS topic groups. The focus of these reviews will be, respectively, to (i) identify the un-resolved methodological challenges and issues that will require systematic investigation by analytical means and/or simulations (e.g. validation or comparisons of alternative methods to address specific challenge(s) relevant for a given TG), and (ii) establish which of the relevant analytical methods are applied in substantive studies and identify common pitfalls in the analyses of real-life data. A guidance document was circulated and the issues were considered in a plenary session led by Gary Collins. This panel will now incorporate the feedback into a revised document, which will be considered by the Steering Committee.

4.4 Simulation Panel (SP)

Simulations are key to the work of all TGs. They will be essential to (i) validate some of the recently developed methods, (ii) compare how the relative advantages and disadvantages of alternative methods depend on the true underlying data structure, and (iii) at time, illustrate the pitfalls and potentially serious errors induced by conventional methods (frequently used in the applied research), especially when the underlying assumptions are violated. This panel develops and promotes principles for, and provides examples of, best practice for simulation studies. Discussions during the BIRS workshop focused on drawing on the literature to identify the key principles, and find useful published examples, which will help TG members with the design, conduct, analysis and reporting of targeted simulation studies. This will help ensure the conduct of simulation studies is as consistent as possible across TGs (especially when they touch on similar issues), and support the accessibility, transparence, and reproducibility principles, that are fundamental to all STRATOS research activities.

4.5 Data Sets Panel (DP)

This panel coordinates the search for, and criteria for use of, publicly available real-life datasets that will be used by different TGs to illustrate applications of the state-of-the-art methods and/or specific methodological issues or pitfalls. For reproducibility, it was agreed all data sets used in STRATOS publications will be publicly available via the STRATOS website. Meta-data will be used to provide an easy overview of datasets. The DP is developing common protocols for these activities.

4.6 Website Panel (WP)

The STRATOS website is central for both (a) communication within the initiative, and (b) for general research community, as the primary access point for (future) guidance documents, software packages and codes, data, as well as information about individual TGs and the initiative as a whole. During the BIRS workshop, a plenary session reviewed the current website and discussed the strategy for future development and enhancements of the website. Key characteristics for a good website, such as appearance, content, functionality, website usability, and search engine optimization, were discussed. Many important properties are already met but several further extensions are required.

It was also generally agreed to make a start toward integrating social media activities into the STRATOS website. The hashtag #STRATOSi was registered using the healthcare hashtags project (http://www.symplur.com/search/stratosi). Members agreed to link to the STRATOS web pages to their individual home and research pages, which will increase the visibility of the STRATOS website.

4.7 Further panels

The work of the remaining panels, that focus mostly on organizational issues, progressed, and is summarized below:

- Membership Panel: the protocol for processing requests to join the initiative has been agreed by the Steering Group and published on the website. This is essential given increasing visibility of the initiative within the international research community results in a increasing number of researchers who would like to join STRATOS.

- Contact with other Organizations: STRATOS is closely linked with the International Society of Clinical Biostatistics; links are being developed with the International Biometric Society (see section 5 on "post-BIRS activities). A list of other relevant scientific Societies that will be approached was drawn up.

- Knowledge Translation: In collaboration with the Publication Panel (PP), the panel will spearhead publicizing the work of the initiative to the target audiences. A number of imaginative approaches for this emerged during discussions at the BIRS workshop.

- Bibliography Panel: it was agreed this panel will oversee creation of a comprehensive bibliography for the initiative, which will be available on the STRATOS website.

5 STRATOS post-BIRS Activities

Informed and stimulated by the discussions at the BIRS meeting, the following activities have taken place since:

- Invited, videorecorded, STRATOS session at the International Biometric Conference (July 12, 2016, Victoria, Canada; https://biometricconference.org/)

- STRATOS Mini-symposium at ISCB2016 (25 August, 2016, Birmingham, UK; http://www.iscb.info/ISCB2016.html).

- Special STRATOS session at the HEC2016 conference (Health—Exploring Complexity: An Interdisciplinary Systems Approach; September 1, 2016, Munich, Germany; http://www.hec2016.eu/).

Slides of all 19 talks presented at these sessions and mini-symposia are available on the STRATOS website.

6 Future Challenges and Plans

6.1 Big data

The emergence of 'Big Data', and the teams of researchers who focus on this complex topic, 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 many diverse areas and many types of data sources: e.g. 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 in order to respond to specific challenges induced by the type of 'Big Data' relevant to their focus. As the TG reports (summarized in Section 3 above) illustrate, this is already happening. For example, the missing data topic group (TG1) has been involved in development and dissemination of techniques for large electronic health record datasets; the design TG5 is including large routinely collected data examples in their first publication. Furthermore, challenges related to variable selection and modeling of functional forms for continuous variables (that are focus of TG2) are of paramount importance in the analyses of datasets. Finally, the High Dimensional DataTG9 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.

6.2 Future funding

It was agreed that TGs should be on the alert for finding external funds to support future, smaller workshops, which could facilitate pushing forward their research agendas. Such workshops could constitute attractive 'satellite meetings' to established conferences, thereby minimizing travel costs. For example, the UK MRC Methodology Panel funds such workshops.

Steering Group members will also reach out to national funding agencies to explore opportunities for receiving financial support for the vital, but less glamorous, work of the underpinning research panels.

Complementing this, the TGs are encouraged to identify topics for primary research, and—as the BIRS workshop convincingly demonstrated—STRATOS provides a natural forum for the emergence of exciting new primary research collaborations.

6.3 Future Meeting of the entire STRATOS Initiative team

The Steering Group believes that the STRATOS activities will greatly benefit from a second international meeting of all TGs and panels, in 2018, prior to releasing the first of the STRATOS guidance documents on our interactive website.

To this end, immediately after the BIRS meeting, in July 2016 an application was submitted for a research week at the Mathematisches Forschungsinstitut, in Oberwolfach (MFO), Germany, in 2018. Several members of the initiative have participated in previous MFO workshops and know well the stimulating, inspiring and interactive atmosphere at this special place.

Prior to the 2018 meeting, STRATOS topic groups and panels will circulate their draft guidance documents, focusing on the issues fundamental for their respective areas. At the meeting, we will use a structured format to discuss, revise and finalize these documents.

The second key focus of the planned 2018 meeting will be on novel research projects requiring development and validation of new multi-stage, methodology, to be developed jointly by several Topic Groups, to *simultaneously* address analytical challenges frequently encountered in real-life observational studies, and as well as the emerging challenges induced by Big Data. Specifically, the following challenges are envisaged:

- identification of efficient and flexible multivariable methods for variable selection and modeling functional dose-response relationships, both for low and high dimensional data;

- robust, accessible methodology for analyzing complex survival and multi-state data, and

- dynamic approaches to causal inference.

Alongside, and related to this, an important, highly innovative aspect of the meeting will focus on further steps for combining cutting-edge methods from different fields, in order to match the complexity of analyses of real-life observational studies (where many of these analytical challenges have to be *simultaneously* addressed). For example, a running thread will be addressing coherently, across the three topics, the challenges of missing data, unmeasured confounding, and measurement errors, common to most observational studies. Given the complexity of the data and methodologies, while analytic work is key to developing new methods of estimation and statistical inference, this must be complemented by validation and comparison of alternative methods in comprehensive, contextually relevant simulation studies.