STRengthening Analytical Thinking for Observational Studies

Challenges, aims and the general approach

Willi Sauerbrei[1]
Michal Abrahamowicz[2], Marianne Huebner[3] and Ruth Keogh[4]

[1] Institute for Medical Biometry and Statistics, Medical Center University of Freiburg, Germany
[2] Department of Epidemiology and Biostatistics, McGill University, Montreal, Canada
[3] Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA
[4] Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK

http://stratos-initiative.org/

STRATOS INITIATIVE

UNIVERSITÄTS KLINIKUM FREIBURG

STRATOS Initiative: STRengthening Analytical Thinking for Observational Studies

- **The overarching long-term goal:**

  To improve design and statistical analyses of observational studies in practice
  by *'closing the gap'* between (i) recent relevant developments in statistical methodology *versus*
  (ii) methods applied in real-life observational studies

- Specific aims:
  - Develop **evidence-supported guidance** for statistical issues of practical importance (*through discussions among experts with different views, and simulations to systematically assess and compare alternative methods*)
  - Provide **guidance at several levels** of statistical knowledge
  - Start with **state-of-the-art** guidance for issues where there is consensus and necessary evidence
  - **Identify and explore complex analytical challenges requiring more primary research** and/or **combining expertise** in different areas of statistical research

Different levels of statistical knowledge

Level 1: Low statistical knowledge

- Many studies are analysed by researchers with this level

Level 2: Experienced statistician

- Uses methodology which is ok, however often not according to state of the art

Level 3: Expert in a specific area

- Performs research to improve statistical models and adapts them to complex real problems.

STRATOS – History and Milestones

- **2011 Epi Subcom** at 42th Int Soc Clin Biostatistics (ISCB) in Ottawa
- **2013: Initiative launched** at 44th ISCB in Munich
- **2014: 1st STRATOS paper,** *Statistics in Medicine* 2014
- **2016 & 2019: 2 General meetings**, Banff Int Res Station (BIRS), Canada
- **2021: General virtual meetings**
- **Series in Biometric Bulletin (since 3/2017), to proceed until 4/2024**

As of 2023: >100 members (from 20 countries on 5 continents)

STRATOS Topic Groups (TGs)

Topic Group		Chairs
1	**Missing data**	James Carpenter (UK), Kate Lee (AUS)
2	**Selection of variables and functional forms in multivariable analysis**	Georg Heinze (AUT), Aris Perperoglou (UK), Willi Sauerbrei (GER)
3	**Initial data analysis**	Marianne Huebner (US), Carsten Oliver Schmidt (GER)
4	**Measurement error and misclassification**	Victor Kipnis (US), Pam Shaw (US)
5	**Study design**	Suzanne Cadarette (CAN), Mitchell Gail (US)
6	**Evaluating diagnostic tests and prediction models**	Ewout Steyerberg (NL), Ben van Calster (NL)
7	**Causal inference**	Els Goetghebeur (BEL), Ingeborg Waernbaum (SWE)
8	**Survival analysis**	Michal Abrahamowicz (CAN), Malka Gorfine (IS), Terry Therneau (US)
9	**High-dimensional data**	Lisa McShane (US), Joerg Rahnenfuehrer (GER), Riccardo de Bin (NOR)

Chairs from 11 countries and 4 continents

STRATOS cross-cutting Panels

Panel		Chairs and Co-Chairs	
MP	**Membership**	Chairs:	James Carpenter (UK), Willi Sauerbrei (GER)
PP	**Publications**	Chairs:	Bianca De Stavola (UK), Mitchell Gail (US), Pamela Shaw (US), Mark Baillie (CH)
GP	**Glossary**	Chairs:	Martin Boeker (GER), Marianne Huebner (US)
WP	**Website**	Chairs:	Joerg Rahnenfuehrer (GER), Willi Sauerbrei (GER)
RP	**Literature Review**	Chairs:	Gary Collins (UK), Carl Moons (NL)
BP	**Bibliography**	Chairs:	to be determined
SP	**Simulation Studies**	Chairs:	Michal Abrahamowicz (CAN), Anne-Laure Boulesteix (GER)
DP	**Data Sets**	Chairs:	Saskia Le Cessie (NL), Maarten van Smeden (NL)
TP	**Knowledge Translation**	Chair:	Maarten van Smeden (NL)
CP	**Contact Organisations**	Chairs:	Willi Sauerbrei (GER)
VP	**Visualisation**	Chairs:	Mark Baillie (CH)
OS	**Open Science**	Chair:	Sabine Hoffmann (GER)

Example of a Challenging Observational Study:
Hydrochlorothiazide use vs. Non-Melanoma Skin Cancer (NMSC)

Background:

- Hydrochlorothiazide (HCTZ) is a popular antihypertensive drug, known to increase the sensitivity of the skin to sunlight and UV radiation
- UV exposure is an important risk factor for NMSC, the most common cancer worldwide
- Emerging evidence of NMSC risk associated with cumulative HCTZ exposure

Objective:

To Respond to Health Canada (federal Ministry of Health) Query:
If and How NMSC risk increases with Cumulative Duration of HCTZ use?

Many challenges

- Challenges at Intersection of: Design (TG5), Survival Analysis (TG8) and Causal Inference (TG7)
- Causal Inference (TG7): DAG to identify Unmeasured Confounders for HCTZ → NMSC association
- Time-Varying Exposure metric: Intersection of Design (TG5) & Causal Inference (TG7) & Survival (TG8)
- Initial Data Analysis (TG3) & Visualisation panel
- Exposure: Measurement error & misclassification (TG4)
- Errors in Cumulative Exposure due to Interval Censoring of the event times
- Outcome & Modeling: Survival Analysis (TG8)
- Covariates (selection & modeling): Selection of variables and functional forms (TG2)
- Imputing Unmeasured Confounders: Intersection of Missing Data (TG1) & Causal Inference (TG7) & Survival (TG8)
- Further Analytical Challenges: Evaluating Prediction Models (TG6), Causal Inference (TG7) & Survival (TG8)

Example of a collaboration

- Developing international standards in the analysis of patient reported outcomes (PRO) in cancer clinical trials: methodological issues and STRATOS engagement in the European IMI-SISAQOL project

  talk le Cessie at ISCB 2022

- Setting International Standards in Analysing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials

- PRO data (eq. Quality of Live) pose many methodological issues

Stakeholders involved in SISAQOL-IMI

- Academia
- Industry
- Regulators (including EMA and FDA `representatives')
- Health technology assessment bodies
- Clinicians
- Methodological and applied statisticians
- PRO experts
- Patient representatives

- And STRATOS

STRATOS guidance for analysts with limited statistical knowledge

See overview in Biometrical Bulletin article (1/2023)
Heinze et al.

new Level 1 page, still password protected

Conclusions

- Observational studies pose several analytical challenges
- Some frequently encountered challenges require **combining expertise from different areas of statistical research**
- For some issues, there are **several alternative statistical approaches** but little solid evidence re:
  i. **Which method(s) work best?**
  ii. How their **relative performance depends on data structure?**

  So there is **no** *state-of-the-art,* **further simulation studies** may be useful

- Other **complex issues require new analytical developments**
- Many of these *state-of-the-art* issues are addressed in recent papers by STRATOS Topic Groups
- Future STRATOS **guidance for data analysts with limited statistical background** will focus on (i) **choice of appropriate easy-to-implement methods** and (ii) **limitations of some popular approaches**

Next meetings

- **August 31, 2023** Full day Mini-Symposium at **ISCB 44**, Milan, Italy
- **September 7, 2023** Satellite Symposium at the **CEN-Conference**, Basel, Switzerland

- Presentations at these two meetings could be the basis for an overview paper/series of papers on behalf of the STRATOS initiative

- WHO wants to organize STRATOS sessions or meetings at other conferences?

2.00 – 3.20	**Session 1**
2.00 – 2.20	Sauerbrei for ExCom, recent developments and short-term goals
2.20 – 2.40	TGX1 – Gorfine for TG8
2.40 – 3.00	Talk 1-TG – Proust-Lima: TG4/8 project
3.00 – 3.20	Panel 1 – Boulesteix for Simu-Panel
3.30 – 4.40	**Session 2**
3.30 – 3.50	TGX2 – Shaw/Kipnis for TG4
3.50 – 4.10	TGX3 – Huebner for TG3
4.10 – 4.40	Project 1 – Carpenter p-values
5.00 – 6.00	**Session 3**
5.00 – 5.20	TGX4 – De Bin for TG9
5.20 – 5.40	Talk 2-TG – Baillie: TG2/3 project
5.40 – 6.00	Panel 2 – Boeker for Glossary
6.05 – 7.05	**Session 4**
6.05 – 6.25	TGX5
6.25 – 6.45	Project 2 – Abrahamowicz for Simulation Panel
6.45 – 7.05	Project 3 (if needed, otherwise end earlier)