

# Guidance for key issues of design and analysis of observational studies

## General overview of the aims and strategies of the international initiative

Willi Sauerbrei (Freiburg, Germany)  
for the steering group of the initiative

# Why do we need such an initiative?

---

- Weaknesses of many analyses
- Many analyses are conducted by people with limited statistical knowledge
- Analysis of observational studies – issues are very similar in all areas of science
- Rapid developments of statistical methodology requires guidance and education

# Guidance and education required

---

- Statistical models are always a simplification of real life processes. To improve these models, researchers develop new and more complicated approaches
- Different and partly conflicting approaches are proposed
- Expert knowledge is required to use methods
- Statistical software has to be available

# Current situation

---

- Statistical methodology has seen some substantial development
- Computer facilities can be viewed as the cornerstone
- Possible to assess properties and compare complex model building strategies using simulation studies
- Resampling and Bayesian methods allow investigations that were impossible two decades ago
- Machine learning approaches may be interesting alternatives to more traditional approaches
- Wealth of new statistical software packages allow a rapid implementation and verification of new statistical ideas

**Unfortunately, many sensible improvements are ignored in practical statistical analyses**

# Reasons that improved strategies are ignored

---

- Overwhelming concern with **theoretical aspects**
- Very **limited guidance** on key issues that are **vital in practice**, discourages analysts from utilizing more sophisticated and possibly more appropriate methods in their analyses

# Improvement

---

At least **two tasks** are essential

1. **Experts** in specific methodological areas have to work towards **developing guidance documents**
2. An ever-increasing need for **continuing education** at all stages of the career

For busy applied researchers it is often **difficult to follow methodological progress** even in their principal application area

- Reasons are diverse
- Consequence is that analyses may be deficient
- **Knowledge** gained through research on statistical methodology needs to be **transferred** to the broader community
- Many **analysts** would be **grateful for** an overview on the current **state of the art** and for **practical expert guidance**

# Aims of the initiative

---

- **Provide guidance documents** for highly relevant issues
- As the statistical **knowledge** of the analyst **varies** substantially, guidance has to keep this background in mind. **Guidance** documents have to be provided **at several levels**
- For the **start** we will concentrate on **state-of-the-art** documents for experienced statisticians (**level 2**)
- Help to identify questions **requiring more primary research**

**The overarching long-term aim is to improve key parts of statistical analyses of observational studies in practice**

# Different levels of statistical knowledge

---

## Level 1: Low statistical knowledge

- Most analyses are done by analysts at that level
- **Point out weaknesses** of approaches often **used despite of problems** (e.g. categorizing continuous variables in the analysis; complete case analysis if a case has missing values in one or more variables)
- Propose **methods** which may not be optimal or state of the art, but which are easy to use and which are **still acceptable** from a methodological point of view
- **Software** should be generally available



# Different levels of statistical knowledge

---

## Level 2: Experienced statistician

- Methodology perhaps slightly below **state of the art**, but doable by every experienced analyst
- **Advantages and disadvantages** of competing approaches, point to the importance and implications of underlying assumptions
- Sufficient guidance about **software** plays a key role that the approaches are also used in practice

# Different levels of statistical knowledge

---

## Level 3: Expert in a specific area

- To improve statistical models and to adapt them to complex real problems, researches **develop** new and **more complicated approaches**
- Advantages and usefulness in practice are unclear
- Often, advantages are presented in a small number of examples and in specific situations but a **more systematic comparison to the state of the art is needed**
- **Software** requires **specific knowledge** and may not be generally available
- **Overview of recent research** with statements about possible advantages and disadvantages is needed
- Could help to **identify** important **weaknesses** of **level 2** proposals
- Help to identify **areas** needing **more** methodological **research**
- Trigger the development of **software** for more general use

# Short term aims

---

- For a small number of **highly relevant topics** we will try to assess the current **state of practice** and identify current documents which try provide some guidance
- Try to find '**agreement**' **what to recommend** within each of the three levels mentioned
- Documents have to be **understood** and at least broadly **accepted**
- **Positive examples** from the current literature will help to reach the latter goal

# Long term aims

---

- **Improvement of statistical analyses**, acceptance of guidance documents by analysts could be a cornerstone
- The percentage of analysts who reflect only the '**level 1**' knowledge should **decrease substantially**
- Guidance **documents** have to be **regularly improved**. Based on evidence some approaches should be ready to be moved from level 3 into a level 2 recommendation
- **Software** is generally available and usable at a broader level
- Number of **topics** is large and there is often a **relation between them** or one is influenced by the other.  
For example, variable selection and missing data: First it is important to **derive guidance for both of them separately**. For practical reasons it is highly relevant to **consider the implications of** missing data **on** guidance for variable selection.

# Program

---

## **Design and analysis of observational studies in medical research**

<b>Willi Sauerbrei</b>	General overview of the aims and strategies of the international initiative
<b>Doug Altman</b>	Why we need guidance documents for the design and analysis of observational studies
<b>James Carpenter</b>	Guidance initiative: a route map, and the journeying of the missing data topic group (TG 1) Discussion 1
<b>Coffee break</b>	
<b>Saskia Le Cessie</b>	TG 3: Descriptive and initial data analysis
<b>Michal Abrahamowicz</b>	TG 2: Selection of variables and functional form; flexible approaches improve estimation and inference
<b>Els Goetghebeur</b>	TG 7: Causal Inference – at work
<b>Petra Macaskill</b>	TG 6: Evaluating diagnostic tests and prediction models
<b>Helmut Küchenhoff</b>	TG 4: Measurement error
<b>Doug Altman</b>	TG 5: Design  Discussion 2

# Steering Group

---

## **Members:**

- Michal Abrahamowicz (Mc Gill, Montreal, Canada)
- Doug Altman (Oxford, UK)
- James Carpenter (London, UK)
- Ray Carroll (College Station, US)
- Stephen Evans (London, UK)
- Mitch Gail (NCI, Bethesda, US)
- Els Goetghebeur (Gent, Belgium)
- Göran Kauermann (Munich, Germany)
- Saskia Le Cessie (Leiden, The Netherlands)
- Petra Macaskill (Sydney, Australia)
- Jörg Rahnenführer (Dortmund, Germany)
- Willi Sauerbrei (Freiburg, Germany)
- Sebastian Schneeweiss (Harvard, Boston, US)
- Ewout Steyerberg (Rotterdam, The Netherlands)
- Andrew Vickers (Sloan Kettering, New York, US)

# Aim of this symposium

---

Informing you!  
Getting you involved!

- Interest further colleagues to work on guidance documents  
For both the seven topics presented and on some new topics, to be proposed to the Steering Group
- Afternoon session of the members of the initiative  
Interested colleagues are very welcome!

For Further information

Willi Sauerbrei

(IMBI, University Medical Center Freiburg, Germany,  
wfs@imbi.uni-freiburg.de)