## STRengthening Analytical Thinking for Observational Studies: the STRATOS initiative

#### **Catherine Quantin**

University of Burgundy, University of Paris-Saclay, France
University of Liège, Belgium

#### & Willi Sauerbrei

**University of Freiburg, Germany** 

for the STRATOS Initiative

http://stratos-initiative.org/



12<sup>th</sup> European Public Health Conference 20 – 23 November 2019, Marseille, France

# The STRATOS initiative Current situation in statistical methodology

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

## **STATA** new procedures in 2018

#### ERM=Endogeneity +Selection +Treatment

Combine endogenous covariates, sample selection, and endogenous treatment in models for continuous, binary, ordered, and censored outcomes.

Take your causal inference to a whole new level.

Learn more »



#### Latent class analysis (LCA)



Discover and understand the unobserved groupings in your data. Use LCA's model-based classification to find out

- how many groups you have,
- who is in those groups, and
- what makes those groups distinct.

earn more »

#### bayes: logistic ... and 44 more

Continuous
Binary
Categorical Ordinal
Multilevel models
Censoring GLM
Truncation
Sample selection
Panel data Count
Zero-inflated Supplies

Type bayes: in front of any of 45 Stata estimation commands to fit a Bayesian regression model.

earn more.

#### Markdown & dynamic documents

#### Get this



- Create webpages from Stata
- Intermix text, regressions, results, graphs, etc.
- See changes in data or commands automatically reflected on webpage

l oarn more v

#### **Linearized DSGEs**

$$\begin{array}{rcl} p_t &=& \beta E_t(p_{t+1}) + \kappa x_t \\ x_t &=& E_t(x_{t+1}) - (r_t - E_t(p_{t+1}) - z_t) \\ r_t &=& \psi p_t + u_t \\ u_{t+1} &=& \rho_w u_t + \epsilon_{t+1} \end{array}$$

Write your model in simple algebraic form. Stata does the rest: solve model, estimate parameters, estimate policy and transition matrices (with Cls), estimate and graph IRFs, and perform forecasts.

 $z_{t+1} = \rho_z z_t + \xi_{t+1}$ 

Learn more »

#### Finite mixture models (FMMs)



- 17 estimators and combinations
- Continuous, binary, count, ordinal, categorical, censored, and truncated outcomes
- Survival outcomes

Learn more

#### Spatial autoregressive models

#### Because

sometimes

where you are

matters.

Learn more »

#### Interval-censored survival models



Fit any of Stata's six parametric survival models to interval-censored data. All the usual survival features are supported: stratified estimation, robust and clustered SEs, survey data, graphs, and more.

earn more »

#### Nonlinear multilevel mixed-effects models



When ... your science ... says ... your model ...

nonlinear in its parameters

Learn more »

#### Mixed logit models: Advanced choice modeling

Do you walk to work, ride a bus, or drive your car? Which of three insurance plans do you buy? Which political party do you vote for?

We make dozens of choices every day. Researchers have access to gaggles of data about those choices. Mixed logit introduces random effects into choice modeling and thereby relaxes the IIA assumption and increases model flexibility.

Learn more ×

#### Nonparametric regression



When you know something matters.

But have no idea how.

Learn more

#### Create Word documents from Stata

- Automate your reports
- Write paragraphs and tables to Word documents
   Embed Stata results and graphs in paragraphs and tables
- Customize formatting of text, tables, and cells
- 0001011120101110101

Learn more »

Create PDFs, too!

#### Bayesian multilevel models



Small number of groups? Many hierarchical levels? Prefer making probability statements?

Consider Bayesian multilevel modeling

Learn more »

#### Threshold regression



Your time-series regression may change parameters at some point in time or at multiple points in time. The activity of foraging aimlass light follow a completely different pattern at temperatures above some threshold. You may not know the value of that threshold. Finding such thresholds and estimating the parameters within the reqiness is what threshold regression does.

Learn more »

#### Panel-data tobit with random coefficients



Stata has long had estimators for random effects (random intercepts) in panel data.

Now you can have random coefficients, too.

Learn more »

#### Search, browse, and import FRED data



#### Multilevel regression for interval-measured outcomes

Incomes are sometimes recorded in groupings, as are people's weights, insect counts, grade-point averages, and hundreds of other measures. Often

#### Multilevel tobit regression for censored outcomes

Left-censoring, right-censoring, both

#### Panel-data cointegration tests



#### Tests for multiple breaks in time series



## **Current situation in practical analyses**

 Unfortunately, many sensible improvements are ignored, such as the use of spline functions in regression analyses

## Reasons why 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

## How should medical science change?

 Many more gains are possible if the waste and inefficiency in the ways that biomedical research is chosen, designed, done, analysed, regulated, managed, disseminated, and reported can be addressed." The Lancet Research, Macleod et al., 2014

## Better use of statistical methods

- At least two tasks are essential:
  - 1. Experts in specific methodological areas have to work towards developing guidance
  - 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 are often 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 guidance

## **Aims of the STRATOS Initiative**

- Provide evidence supported guidance for highly relevant issues in the design and analysis of observational studies
- For the **start** we will concentrate on **state-of-the-art** guidance and the necessary evidence
- As the statistical knowledge of the analyst varies substantially, guidance has to keep this background in mind. Guidance has to be provided at several levels (Low statistical knowledge, Experienced statistician, Expert in a specific area)

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

## **Organisational Structure**

- STRATOS is led by an **Executive Committee** and a **Steering Group**.
- There are nine **Topic Groups** and eleven cross-cutting **Panels**.

## **Topic groups**

Topic Group		Chairs	
1	Missing data	James Carpenter, Kate Lee	
2	Selection of variables and functional forms in multivariable analysis	Georg Heinze, 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, Suzanne Cadarette	
6	Evaluating diagnostic tests and prediction models	Gary Collins, Carl Moons, Ewout Steyerberg	
7	Causal inference	Els Goetghebeur, Ingeborg Waernbaum	
8	Survival analysis	Michal Abrahamowicz, Per Kragh Andersen, Terry Therneau	
9	High-dimensional data	Lisa McShane, Joerg Rahnenfuehrer	

## **Cross-cutting panels**

Panel		Chairs and Co-Chairs		
MP	Membership	Chairs:	James Carpenter, Willi Sauerbrei	
PP	Publications	Chairs:	Bianca De Stavola, Stephen Walter	
••		Co-Chairs:	Mitchell Gail, Petra Macaskill	
GP	Glossary	Chairs:	Simon Day, Marianne Huebner, Jim Slattery	
WP	Website	Chairs:	Joerg Rahnenfuehrer, Willi Sauerbrei	
RP	Literature Review	Chairs:	Gary Collins, Carl Moons	
ВР	Bibliography	Chairs:	to be determined	
SP	Simulation Studies	Chairs:	Michal Abrahamowicz, Anne-Laure Boulesteix	
DP	Data Sets	Chairs:	Saskia Le Cessie, Aris Perperoglou	
ТР	Knowledge Translation	Chair:	Suzanne Cadarette	
		Co-Chair:	Catherine Quantin	
СР	Contact Organisations	Chairs:	Willi Sauerbrei	
VP	Visualisation	Chairs:	Mark Baillie	

## Difficulties to get agreement on existing methods

TG2 – Variable and function selection – different opinions

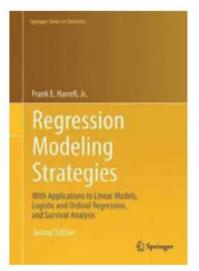


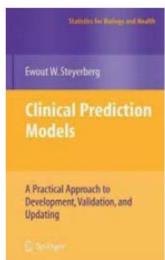


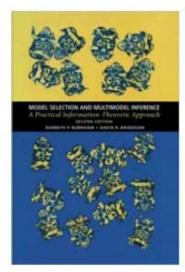
Variable selection

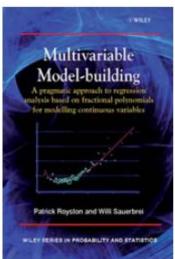












(Harrell, 2001; Steyerberg, 2009; Burnham & Anderson, 2002, Royston & Sauerbrei, 2008)

- → Emphasis on different aims
- → Different philosophies
- → Joint work (except B&A) in STRATOS!

## More new methods needed?

"[...]It becomes more and more difficult to get an overview of existing methods, not to mention the overview of their respective performances in different settings.

[...] Moreover, it is well known that studies comparing a suggested new method to existing methods may be (strongly) biased in favor of the new method.

#### LETTER TO THE EDITOR

**Biometrical Journal** 

On the necessity and design of studies comparing statistical methods

Boulesteix et al for the Simulation Panel of the STRATOS initiative, 2018

# Guidance for whom? Needed by many stakeholders!! analysts with different levels of knowledge, teachers, reviewers, journalists, .....

### Researchers

# First in a Series of Papers for the Biometric Bulletin

STRATOS initiative – Guidance for designing and analyzing observational studies

## **STRATOS**

Willi Sauerbrei<sup>1</sup>, Marianne Huebner<sup>2</sup>, Gary S. Collins<sup>3</sup>, Katherine Lee<sup>4</sup>, Laurence Freedman<sup>5</sup>, Mitchell Gail<sup>6</sup>, Els Goetghebeur<sup>7</sup>, Joerg Rahnenfuehrer<sup>8</sup> and Michal Abrahamowicz<sup>9</sup> on behalf of the STRATOS initiative.

Short papers from all Topic Groups have appeared (TG6 to appear in December issue)

#### **Consumers**

Guidance for designing and analysing observational studies:

The STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative

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

Volume 26 Number 3 | Medical Writing September 2017 | 17

Journal of the European Medical Writers Association (EMWA)

## More Information

 Copies of a STRATOS Poster were deposited at the EUPHA desk; it is also available on the STRATOS website:

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

Thank you for your attention!