



**University of
Zurich** ^{UZH}

Master Program in Biostatistics

Annual Report

Academic Years 2013–2016



Contents

1	Highlights between Fall 2013 and Spring 2016	5
2	Accomplishments and Goals	7
3	Facts and Figures	10
3.1	Recruitment	10
3.2	Theses	11
3.3	Publications from Statistical Consulting Projects or Master Theses	12
3.4	Modules	14
3.5	Talks and Seminars	16
3.6	Faculty and Staff	18
4	Interview with Mark Robinson	21

Preface

This is the second edition of our not regularly spaced report series, therefore it covers the three academic years since the last edition. We produce it as an opportunity to look back and consolidate numbers; showing that the program *is* actually growing and flourishing: enrolled student and available module numbers are up, and through process optimization administrative costs are down.

In the report, we show that we had a small dip in enrollment last fall. Not directly as a consequence but with more focus we were especially carefully tackling the following tasks.

- (i) We have moved our webpage to the new url www.biostat.uzh.ch. In addition a new responsive design allows our web pages to look good on all devices.
- (ii) We have created a profile on facebook for the program www.facebook.com/uzh.biostat
- (iii) We have made minor adjustments to the program and also to the directly linked programs. *Minor in Applied Probability and Statistics* and *PhD Program in Epidemiology and Biostatistics* to better align forces.
- (iv) We are currently working on other PR products for our program.

Keep browsing our web pages for updates of this! I hope you enjoy the report as much we do.

March 2017

Reinhard Furrer
Managing Director

1 Highlights between Fall 2013 and Spring 2016

Graduates

Between Fall 2013 and spring 2016 we proudly saw 14 students graduate from the program. Part of them work now as biostatisticians in research related positions in the public and private sector, part of them started PhD programs either at UZH or elsewhere. We are happy to remark that 5 of them have obtained positions in the closely linked PhD Program in Epidemiology and Biostatistics of UZH. The pictures show the master ceremony of the Faculty of Science in the Spring Semester 2014.



Price of the Faculty of Science for Master Thesis

On December 12, 2014 our former student Lukas Bütkofer, who graduated in spring 2014, received a price from the Faculty of Science for his master thesis, entitled "Influence of ungulate browsing on tree regeneration" and supervised by Torsten Hothorn. The Faculty of Science awards a price for the best 5% of MSc theses in the respective subject area. More information on Lukas and his path can be found below.



R courses

In the fall of 2013 we started to give a one week introduction course to R for our incoming students just before the lectures of the flu semester start. From the beginning this course has been popular also among other students of UZH, priority has always been given to master in biostatistics students. For the fall 2014 edition, we have switched to a new format. The students learn about R in a flipped classroom format with videos produced by Reinhard Furrer, Torsten Hothorn and Owen Petchey (IEU) and face-to-face sessions for practical work.

Starting in spring 2014 we have added a course on Advanced R programming, targeted at advanced master students and PhD students from the PhD Program in Epidemiology and Biostatistics as well as all other interested students meeting the requirement of being proficient R users.



BBQ

Our first Master Program potluck BBQ took place in 2013 and due to weather conditions we had to invade the Furrer household for this occasion. In Spring 2014 (pictures) and 2015 we could celebrate our annual BBQ party in the forrest near Kyburg and in 2016 near Dübendorf. We team up with the PhD Program in Epidemiology and Biostatistics leading to great occasions of sharing food and drinks that our multi-national students and faculty bring to share as well as of many interesting and passionate discussions about only sometimes statistics but more often the world at large. We are proud to say that our students are an extraordinary group of individuals!

2 Accomplishments and Goals

In the period between Fall 2013 and Spring 2016 we have achieved the following points:

- A steady increase of applications from and recruitment of UZH and Swiss students can be observed, see figures below. Hopefully, the low numbers in spring 2016 (and fall 2016) are simply a temporary drawback.
- Moreover, there is a clear and steady increase in students following our lectures, see figures below.
- The Master Program in Biostatistics has proven to be a source for recruiting good Phd students for our in-house program.
- The structure of the program that was implemented initially has seen only one major change (STA425 Survival Analysis become compulsory in Spring 2015). With experience we can confirm that the initial emphasis on compulsory basic course and the increasing degree of independence of students' work while advancing in the program has been the right choice to successfully train knowledgeable and competent young statisticians.
- The Master Program in Biostatistics is increasingly viewed as not only being part of the Faculty of Science but also of EBPI (Epidemiology, Biostatistics and Prevention Institute). The major reasons for this is the participation of many of the master students in the lecture "EPI301 Introduction to Epidemiology" by M. Puhon and colleagues. Moreover, the number of

projects in the consulting module that are initiated by EBPI researchers is increasing.

- One of our goals that we did not accomplish is to be successful with an application at the European Union for a summer program in infectious disease epidemiology together with the University of Ulm, Germany, Hasselt University, Belgium and Stockholm University, Sweden. Our submitted application spring 2015 has been refused funding, potentially also due to the special status of Switzerland after the mass immigration vote in spring 2014. In the spring of 2016 we decided to organise a summer school for students from Ulm and Zurich bilaterally, the first edition will take place in September 2017 in Zurich, locally funded by the Faculty of Science and Ulm University.

We aim to provide a high-quality and pertinent training to future biostatisticians. In order to judge if we indeed accomplished this we have asked our graduate Lukas Bütikofer, mentioned already above, what he is doing now and how he judges how well he was prepared for his job by the program.

Lukas works as a senior biostatistician at the CTU in Bern, there his duties include statistical analysis of observational studies, clinical trials and systematic reviews/meta-analyses, statistical and methodological consulting for clinical investigators, and contribution to study protocols, final reports and publications of research projects. Lukas

says that he was very well prepared for his job by the program, even if we did not teach him all methods he uses now we prepared him to be able to learn new methods fast himself. Comparing his with the background of his colleagues in Bern, he says that he has an advantage since he got a solid technical/mathematical background while still being literate in the clinical and epidemiological applications.

In hindsight, he says, that he should have gotten more in-depth knowledge in survival analysis and in mixed models, areas that we put more focus on since Lukas' graduation, as well as non-parametric and resampling methods, which indeed are subjects we only touch very lightly. Lukas's advice to current students is to practice a lot of programming, since time to dwell on the optimization of implementations once on the jobs is very limited.

Looking forward in time, our goals for the mid-term future are:

- to keep stabilizing student numbers,
- to continue to offer attractive lectures to a larger audience at the Faculty of Science and beyond
- to implement our collaboration with Ulm in the Zurich-Ulm Biostatistics Summer School (ZUBSS)



3 Facts and Figures

3.1 Recruitment

Applications	Total	Admitted	Recruited	Dropouts
Fall 2013	16	12	3	2
Spring 2014	10	6	3	1
Fall 2014	15	11	3	1
Spring 2015	16	11	8	1
Fall 2015	18	14	10	1
Spring 2016	5	3	2	NA
Total	85	57	29	6

Applications	Total	Asia	Africa	Australia	North America	South America	Europe	Switzerland
Fall 2013	16	3	3	0	1	0	5	4
Spring 2014	10	2	2	0	2	0	1	3
Fall 2014	15	3	2	0	1	0	4	5
Spring 2015	16	2	3	0	2	0	3	6
Fall 2015	18	0	2	0	3	0	2	10
Spring 2016	5	1	0	0	1	0	3	0
Total	85	11	12	0	10	0	18	28

Applications	Total	Female	Math/ Stats	Life Science
Fall 2013	16	7	10	4
Spring 2014	10	3	3	5
Fall 2014	15	11	9	2
Spring 2015	16	5	9	3
Fall 2015	18	6	4	7
Spring 2016	5	2	2	2
Total	85	34	37	21

3.2 Theses

Student	Title	Date
Maria Sereina Graber	Phylogenetic comparative methods for discrete re- sponses in evolutionary biology (Furrer)	2013
Isaac Gravestock	Bayesian Tree Models, Priors and Posterior Approxi- mations (Held)	2014
Ruizhu Huang	Evaluation of CD4 and CD8 as progression markers for untreated and treated HIV-I infection (Held)	2014
Lukas Bütikofer	Influence of ungulate browsing on tree regeneration (Hothorn)	2014
Bledar Baholli	Reproduzierbarkeit von Überlebenszeitanalysen mit- tels rekursivem Partitionieren (Hothorn)	2014
Katarina Matthes	A comparison of count-based and assembly-based methods for differential splice detection (Robinson)	2014
Sarah Thommen	Statistical monitoring of condemnation rates from Swiss slaughterhouses (Held)	2014
Jipcy Carolina Amador Sulbaran	Disease Progression and Survival in ALS Patients (Hothorn)	2014
Sabine Güsewell	Phenological responses to changing temperatures: representativeness and precision of results from the Swiss Phenological Network (Furrer)	2014
Monika Hebeisen	Estimation from Interval-censored Time-to-event Data: Method Comparison by Simulation based on Gallium Study for Follicular Lymphoma (Hothorn)	2014
Gabrielle Elaine Moser	Spatial Aspects of Forest Monitoring Data and Surface Estimation: An Analysis of the Swiss NFI (Furrer)	2014
Carlos Ochoa Pereira	Novel semiparametric estimation method for the anal- ysis of zero-inflated data: an application to the young forest records of the Swiss NFI 3 (Furrer)	2015
Gilles Kratzer	Simulation based inference in epidemic models (Held)	2015
Burak Kürsad Günhan	Network meta-analysis with integrated nested Laplace approximations (Held)	2016

Student	Title	Date
Muriel Lynnea Buri	Parametric Bootstrap Inference for Transformation Models (Hothorn)	2016
Yuchen Ling	Meta-Analytic-Predictive Priors Using Historical Control Information with the Integrated Nested Laplace Approximation (Held)	2016
Zhongxing Zhang	Residual Analysis for Longitudinal Data (Held)	2016

3.3 Publications from Statistical Consulting Projects or Master Theses

Student	Module	Publication
Rhuizhu Huang	Consulting Spring 2013	Accepted by European Journal of Heart Failure. Prognostic value of long-term blood pressure changes in patients with chronic heart failure.
Katharina Matthes	Master Spring 2014	Genome Biology 2016, 17:12, doi:10.1186/s13059-015-0862-3. Isoform prefiltering improves performance of count-based methods for analysis of differential transcript usage.
Carlos Ochoa	Consulting Spring 2014	Journal of Antimicrobial Chemotherapy 2016, 71(1), 141–151. Relative contribution of biological variation and technical variables to zone diameter variations of disc diffusion susceptibility testing.
Christian Müller	Consulting Spring 2014	PLOS ONE 2015 Oct 30, 10(10): e0140874. doi: 10.1371/journal.pone.0140874, Predictors of High Profit and High Deficit Outliers under SwissDRG of a Tertiary Care Center.
Sarah Thommen	Master Spring 2014	Epidemiology and Infection, 2015, 143, 3423–3433, doi:10.1017/S0950268815000989. A simulation study on the statistical monitoring of condemnation rates from slaughterhouses for syndromic surveillance: an evaluation based on Swiss data.

Student	Module	Publication
Seraina Graber	Consulting Fall 2013	Journal of Infection, 2014, 79(3), 255–263. doi: 10.1016/j.jinf.2014.10.005, Shared Microbiome in Gums and the Lung in an Outpatient Population.
Florian Gerber	Master Spring 2013	Journal of Statistical Software, 2015, 63, DOI: 10.18637/jss.v063.c01, Pitfalls in the Implementation of Bayesian Hierarchical Modeling of Areal Count Data: An Illustration Using BYM and Leroux Models.
Lukas Bütikofer	Consulting Spring 2013	Dental Materials 31(2), e33–e50, Two regression methods for estimation of a two-parameter Weibull distribution for reliability of dental materials.
Isaac Gravestock	Consulting Spring 2013	Journal of Hand Surgery, 2014, 39(7), 1344–1350, DOI: 10.1016/j.jhssa.2014.03.025, Primary Flexor Tendon Repair in Zones 1 and 2: Early Passive Mobilization Versus Controlled Active Motion.
Florian Gerber	Consulting Spring 2012	Anal. Chem., 2013, 85 (21), 10249–10254, DOI: 10.1021/ac402018e, Multiorder Correction Algorithms to Remove Image Distortions from Mass Spectrometry Imaging Data Sets.
Sabine Güsewell	Consulting Spring 2012	Interactive Cardiovascular and Thoracic Surgery, 2014, 19(5), 749–55. doi: 10.1093/icvts/ivu246, Does body mass index impact the early outcome of surgical revascularization? A comparison between off-pump and on-pump coronary artery bypass grafting.

3.4 Modules

Fall semester 2013

Number	Title	Enrollment
STA 260	Practical Introduction to the Statistical Computing Environment R (Hothorn)	20
STA402	Likelihood inference (Furrer)	24
STA404	Statistical Methods in Clinical Research (Held)	9
STA406	Generalized Regression (Hothorn)	8
STA426	Statistical Methods for the Analysis of Microarray and Short-read Sequencing Data (Robinson/Rehauer)	5

Spring semester 2014

Number	Title	Enrollment
STA408	Statistical Methods in Epidemiology (Held)	7
STA425	Survival Analysis (Hothorn)	10
STA480	Biostatistical Journal Club (Furrer)	2
STA490	Statistical Consulting (Furrer/Held)	4
STA660	Advanced R Programming (Hothorn)	12

Fall semester 2014

Number	Title	Enrollment
STA121	Statistische Modellierung (Furrer)	6
STA 260	Practical Introduction to the Statistical Computing Environment R (Seibold)	22
STA291	Statistics and Probability (reading class: Furrer)	1
STA402	Likelihood inference (Furrer)	21
STA404	Statistical Methods in Clinical Research (Held)	14
STA406	Generalized Regression (Hothorn)	12
STA426	Statistical Methods for the Analysis of Microarray and Short-read Sequencing Data (Robinson/Rehauer)	1

Spring semester 2015

Number	Title		Enrollment
STA330	Modeling Dependent Data	(Furrer)	5
STA380	Selected Topics in Statistics	(Furrer)	3
STA408	Statistical Methods in Epidemiology	(Sick)	9
STA422	Bayesian Inference	(Held)	10
STA423	Mixed Models for Correlated Data	(Hothorn)	15
STA425	Survival Analysis	(Hothorn)	12
STA480	Biostatistical Journal Club	(Furrer)	11
STA490	Statistical Consulting	(Held)	5

Fall semester 2015

Number	Title		Enrollment
STA111	Stochastische Modellierung	(Furrer)	6
STA121	Statistische Modellierung	(Furrer)	14
STA 260	Practical Introduction to the Statistical Computing Environment R	(Seibold)	21
STA291	Statistics and Probability	(reading class: Furrer)	3
STA402	Likelihood inference	(Balabdaoui/Held)	36
STA404	Statistical Methods in Clinical Research	(Furrer E/Held)	21
STA406	Generalized Regression	(Hothorn)	16
STA426	Statistical Analysis of High-Throughput Genomic and Transcriptomic Data	(Robinson/Rehrauer)	9
STA428	Bayesian Hierarchical Model	(Roos)	9
STA490	Statistical Consulting	(Held)	3

Spring semester 2016

Number	Title		Enrollment
STA330	Modeling Dependent Data	(Furrer)	9
STA380	Selected Topics in Statistics	(Furrer)	5
STA408	Statistical Methods in Epidemiology	(Sick)	24
STA425	Survival Analysis	(Hothorn)	19
STA451	Selected Topics in Biostatistics	(Nakas)	9
STA480	Biostatistical Journal Club	(Furrer)	14
STA490	Statistical Consulting	(Hothorn)	6
STA660	Advanced R programming	(Hothorn)	21

3.5 Talks and Seminars

Research Seminar in Applied Statistics: non-student talks

Name	Title	Date
M. Pittavino (WHO)	What is...an additive Bayesian Network? Part II	Nov 21 13
C. Chevalier (UZH)	Fast sequential evaluation strategies using Gaussian process models	Jan 9 14
M. Pittavino (WHO)	Rigorous assessment of the replacement rate in disease-free and CAEV-affected goat breedings, via mathematical models	March 28 14
F. Oriani (U Neuchâtel)	Stochastic rainfall simulation: reproducing high-order statistics with the Direct Sampling technique	Apr 10 14
M. Molinaro (UZH)	A Bayesian Spatio-Temporal Geostatistical Model with an Auxiliary Lattice for Large Datasets	Sept 25 17
F. Bachoc (U Vienna)	ML and Cross Validation for covariance function estimation in Gaussian process regression	Nov 6 14
F. Balabdaoui (CEREMADE)	Estimation of a log-concave probability sequence via Maximum Likelihood: Asymptotics and related applications	Nov 12 14
R. Furrer (UZH)	To Kronecker or not to Kronecker, that is the question	Feb19 15
C. Chevalier (UZH)	The very basics of computer experiment: Bayesian optimization and inversion	Feb 26 15
M. Molinaro (UZH)	Parametrization and computational challenges in multivariate Gaussian Markov random fields modeling	March 19 15
M. Pittavino (WHO)	Additive Bayesian networks: the choice of a suitable parameters prior	March 26 15
L. Pasanen (U Oulu)	Bayesian Scale space for images, time series and correlation analysis	May 7 15
D. Nychka (UCAR)	Extremes in Regional Climate: What to do with 8000 Histograms?	March 15 16

Zükost: talks initiated by faculty members of the program

Name	Title	Date
J. Copas (U Warwick)	Correcting for outcome reporting bias in meta-analysis	Oct 31 13
J. Wakefield (U Washington)	The Modeling of Pathogen-Specific Counts for Hand, Foot and Mouth Disease	Dec 5 13
D. Rossell (U Warwick)	RNA-seq and alternative splicing. A high-dimensional estimation, model selection and experimental design problem	March 6 14
F. Bretz (Novartis)	Adaptive Methods in Clinical Trials	Oct 16 14
M. Zucknick (TU Dortmund)	Bayesian models for risk prediction with high-dimensional (integrative) genomics	Oct 30 14
M. Schumacher (U Freiburg)	From conditional survival to dynamic predictions aspects of application, statistical modelling and assessment	Dec 4 14
F. Leisch (U Wien)	Flexible Implementation of Resampling Schemes for Cluster Validation	Apr 15 15
D. Kuonen (Statoo Cons)	A Statistician's "Big Tent" View on Big Data and Data Science	Apr 23 15
M. Posch (U Wien)	Addressing Multiple Objectives in Clinical Trials: Adaptive Designs and Multiple Testing	Nov 19 15
A. Riebler (Norwegian U)	Bayesian hierarchical models for routine use: What do we need?	May 11 16
S.A. Kovalchik (RAND)	The Past, Present, and Future of Prediction in Professional Tennis	May 25 16

3.6 Faculty and Staff

Faculty

Reinhard Furrer	Department of Mathematics, University of Zurich Managing Director
Leonhard Held	Epidemiology, Biostatistics and Prevention Institute, University of Zurich Program Director
Torsten Hothorn	Epidemiology, Biostatistics and Prevention Institute, University of Zurich
Mark Robinson	Institute of Molecular Life Sciences, University of Zurich
Burkhardt Seifert	Epidemiology, Biostatistics and Prevention Institute, University of Zurich

UZH Lecturers

Eva Furrer	Epidemiology, Biostatistics and Prevention Institute, University of Zurich
Hubert Rehrauer	Functional Genomics Center Zurich
Malgorzata Roos	Epidemiology, Biostatistics and Prevention Institute, University of Zurich
Heidi Seibold	Epidemiology, Biostatistics and Prevention Institute, University of Zurich
Beate Sick	Epidemiology, Biostatistics and Prevention Institute, University of Zurich

External Lecturers

Fadoua Balabdaoui	Assistant Professor, Université Paris-Dauphine
Christos Nakas	Research Associate, Inselspital Bern

Staff

Eva Furrer	Epidemiology, Biostatistics and Prevention Institute, University of Zurich Scientific coordinator
Franziska Robmann	Department of Mathematics, University of Zurich

4 Interview with Mark Robinson

RF: You do what is called "statistical bioinformatics". What is that, really?

MARK ROBINSON: I guess the classical statistician would probably call me a "subject-area statistician". So, we apply and develop statistical methods, but for the most part, to large-scale biology experiments only; and when I say large, I typically mean a large number of features measured, but often small in the number of experimental units surveyed. With all the developments in technology, it is really an exciting time to be an applied statistician in the biomedical sciences. While we have specialties in our group (microarrays and sequencing for expression and methylation, single cell measurements), we are interested generally in statistical methods for large-scale datasets. Just as in statistics, the field of bioinformatics is large and diverse, which is also a great feature; there are always multiple ways to attack a problem and new methods to learn.



RF: Your CV, including your citation metrics look impressive. What were your key factors and events for that?



MR: I guess the key events involve moving half way around the world! But, the main factor is, honestly, luck. OK, also some hard work and dedicating time to the tasks at hand played a big role, but ultimately my path was charted with a fair bit of luck. When I finished my M.Sc. in statistics, I knew (or remembered) very little biology. Despite this, Tim (Hughes) took me on as a data analyst in his lab doing microarrays. After a few years of this and a brief sojourn doing proteomics data analysis in a company, it was time to do a PhD and I was lucky to surround myself with great people and great ideas in Terry Speed's and Gordon Smyth's groups at WEHI in Melbourne; to me, WEHI Bioinformatics is the birthplace of statistical bioinformatics. After this, I was again lucky to do a post-doc with Sue Clark in data analy-

sis for epigenome data. And that all lead to establishing my group here in Zurich, where I'm also lucky to have many great colleagues who collaborate with us and send us in new (methodological) directions. I don't have much to say about citation metrics, as I find it all an embarrassing game that academics play. Luckily, this old school thinking towards citation counting as a measure of impact is starting to recede and I hope that modern initiatives, such as DORA (declaration of research assessment), will be embraced. Especially in our field, it is important to look beyond Google Scholar metrics to other important research contributions, such as software (e.g., R packages), training, collaboration and benchmarks.

RF: Which paper are you most proud of and why?

MR: For me, there is no single paper that sticks out. Every primary manuscript is a huge amount of work and they are all special and have their own story. Overall, I am proud of all of them, but probably the "firsts" stick out. My first first-author paper with Tim (FUNSPEC), the first one with Terry (Affy comparison), the first edgeR paper (small-sample negative binomial dispersion estimation), the first postdoc paper (methylation comparison) and the first methodological papers from my Zurich group (BayMeth

or robust edgeR) are a little extra special. I'm also very proud of my current group and that I can now live vicariously through the many "firsts" of my staff and students.



RF: What are the key elements for a successful interdisciplinary collaboration?

MR: We collaborate a lot and I would say that our interactions span the whole scale: sometimes we provide an analysis and never hear anything again, sometimes we hear back 2 years later and a paper is already written and sometimes despite serious effort, we haven't stumbled on that crucial idea yet to push things forward. Of course, sometimes, the data is noisy, biased or a better experiment needs to be designed. I guess the answer depends on your definition of success (you are not asking about citation metrics are you?), but I think the best collaborations come when both parties have a mutual interest and of course, a match on personality helps. For me, the best collaborators are those that have a genuine

interest in the details of the data, as opposed to those who just want a "result".

RF: While some years ago, the quote "In God we trust, all others bring data" was heard everywhere, it seems that now we more and more have to face "Do not trust data that you have not forged yourself". What is the role of statistics/a statistician to regain credibility for quantitative approaches with the larger public but also within scientific publications?

MR: Well, that is indeed a timely question, if you consider for example the recent 'fake news' scandal (although not a new phenomenon). One must also not be fooled by their social media feed; a good chunk of it was probably designed specifically for you by a statistical algorithm. How to regain credibility there? I don't know. In sciences, though, I think we are in a pretty good situation. Most of the good labs in statistical bioinformatics publish not only software and methods, but detailed supporting code to go with it (e.g., to reproduce analyses from raw data). In this way, there are no hidden details and I think a generally open science approach goes a long way towards credibility. One thing I appreciate about the field of statistics is that we generally avoid hype. That probably makes us poor salespeople, but at least the credibility can generally

be counted on.

RF: You and your wife pursued an academic career. What do you talk about at dinner table?

MR: We mostly discuss how to discretely discourage our children from going into academia! They've tagged along to enough conferences now, so I think our plan is working.

Impressum

Design:
Reinhard Furrer

Editor:
Eva Furrer

Photos:
MNF
Marc Latzel
Carlos Ochoa

Address:
University of Zurich
Master Program in Biostatistics
Hirschengraben 84
CH-8001 Zurich

Information:
www.biostat.uzh.ch
masterbiostat@math.uzh.ch

Printed in Switzerland



