VLIR-UOS North South South project
Statistical Research Planning

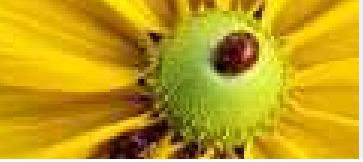
Starting up workshop, UNALM, Lima, Peru

Luc Duchateau, Ghent University

November 7, 2011

1. Overview NSS Statistical Research Planning

- Collaborators
- Main objective
- Experimental design
- Workpackages
- A predecessor: NSS project in Biostatistics
- A contemporary project: Cross cutting initiative in statistics
- VLIR-UOS and statistics



2. NSS partners

- Coordinating universities: Ghent University and Jimma University
- North partners
  - Luc Duchateau, Professor, Faculty of Veterinary Medicine, UGent, coordinator IUC-JU
  - Eddie Schrevens, Professor, Faculty of Bioscience Engineering, KUL, coordinator IUC-UNALM
  - Guido Wyseure, Professor, Faculty of Bioscience Engineering, KUL, coordinator IUC-UCM
- South partners
  - Yehenew Getachew, Lecturer, Faculty of Agronomy, Jimma University, PL IUC-JU
  - Vladimiro Tobar, Professor, Cuenca University,
  - Felipe de Mendiburu, Professor, Faculty of Economy and Planning, UNALM,

3. Main objective

- Encourage and support researchers working under the umbrella of the IUC programme to use proper statistical research planning tools before starting up the actual research in order to
  - Draw valid conclusions from experiments
  - Work more efficiently
- The IUC programme will leave such capacity/ research culture behind at the resp. university

4. Brief history of experimental design

- Rather new discipline, early 20th century
- The godfather is Fisher
- It originates in agronomy
- Few key concepts
  - Randomisation
  - Blocking
  - Nesting
- Analysis (ANOVA) follows

5. Planned activities

- WP1: Enlist experimental designs from IUC projects
- WP2: Develop courses in experimental design
- WP3: Establish website with statistical research planning environment
- WP4: Organise workshops in statistical research planning

5.1. Enlisting experimental design cases

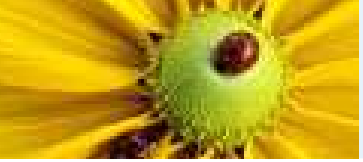
- Enlist statistical research planning problems at the three different IUC programmes
- Present/store/avail such problems in a uniform structure
- Comment, critisise, improve currently used experimental designs in the different IUCs
- Use cases in the coursenotes

5.2. Courses in experimental design

- Develop courses/slides in statistical research planning at different levels
  - Practitioners
  - Statisticians
- These courses could become part of the curricula of the university, e.g., as a prerequisite course for a PhD programme
- Courses available to students through dedicated website
- Courses consist of coursenotes, slides, datasets, R-programmes ...

5.3. Website

- Develop a wesbite that will hold all the material
  - Course material
  - Slides
  - Datasets
  - R programs
  - VIRTEX
  - 1 library of books
- This will serve as a statistical research planning environment

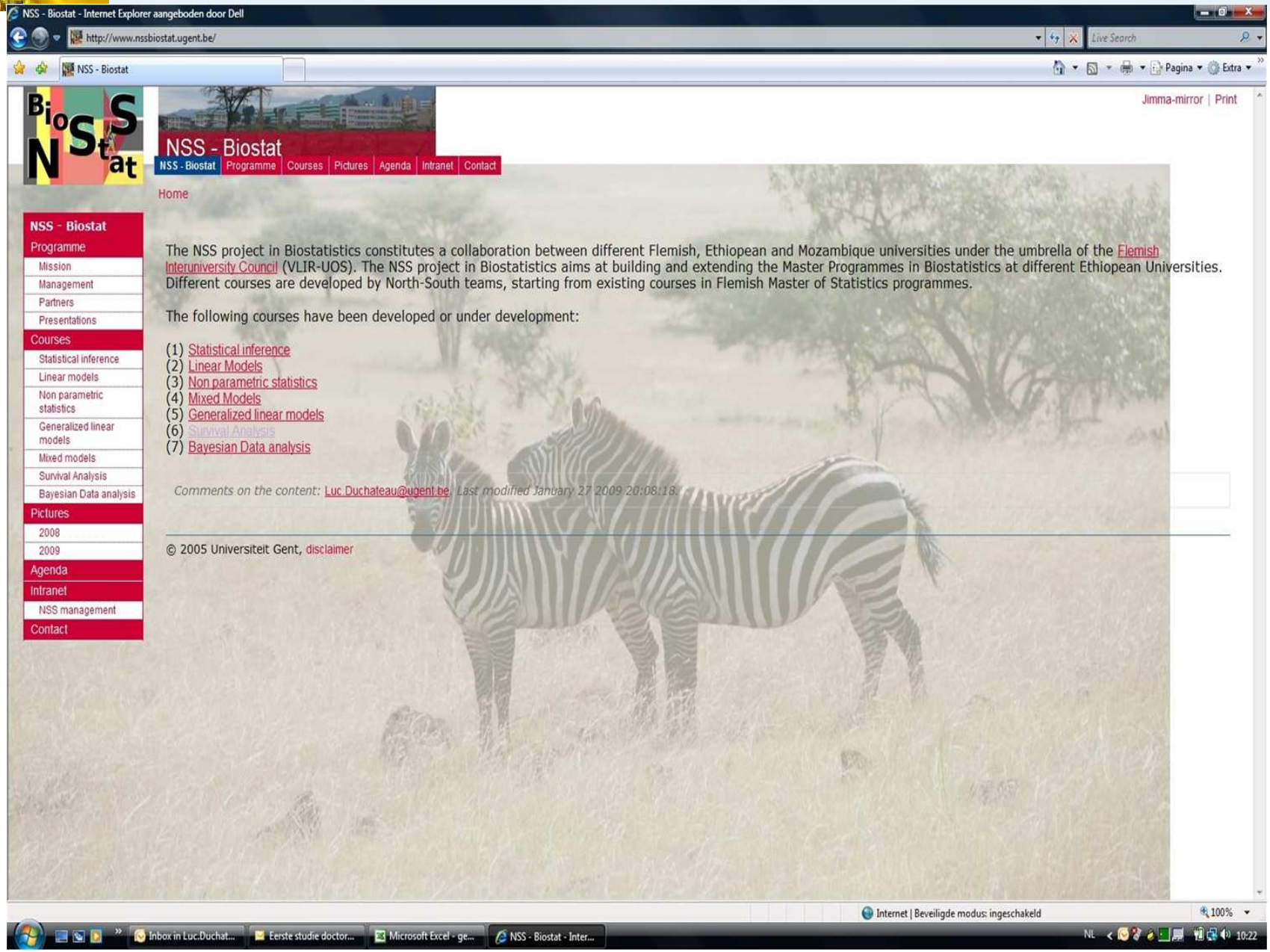


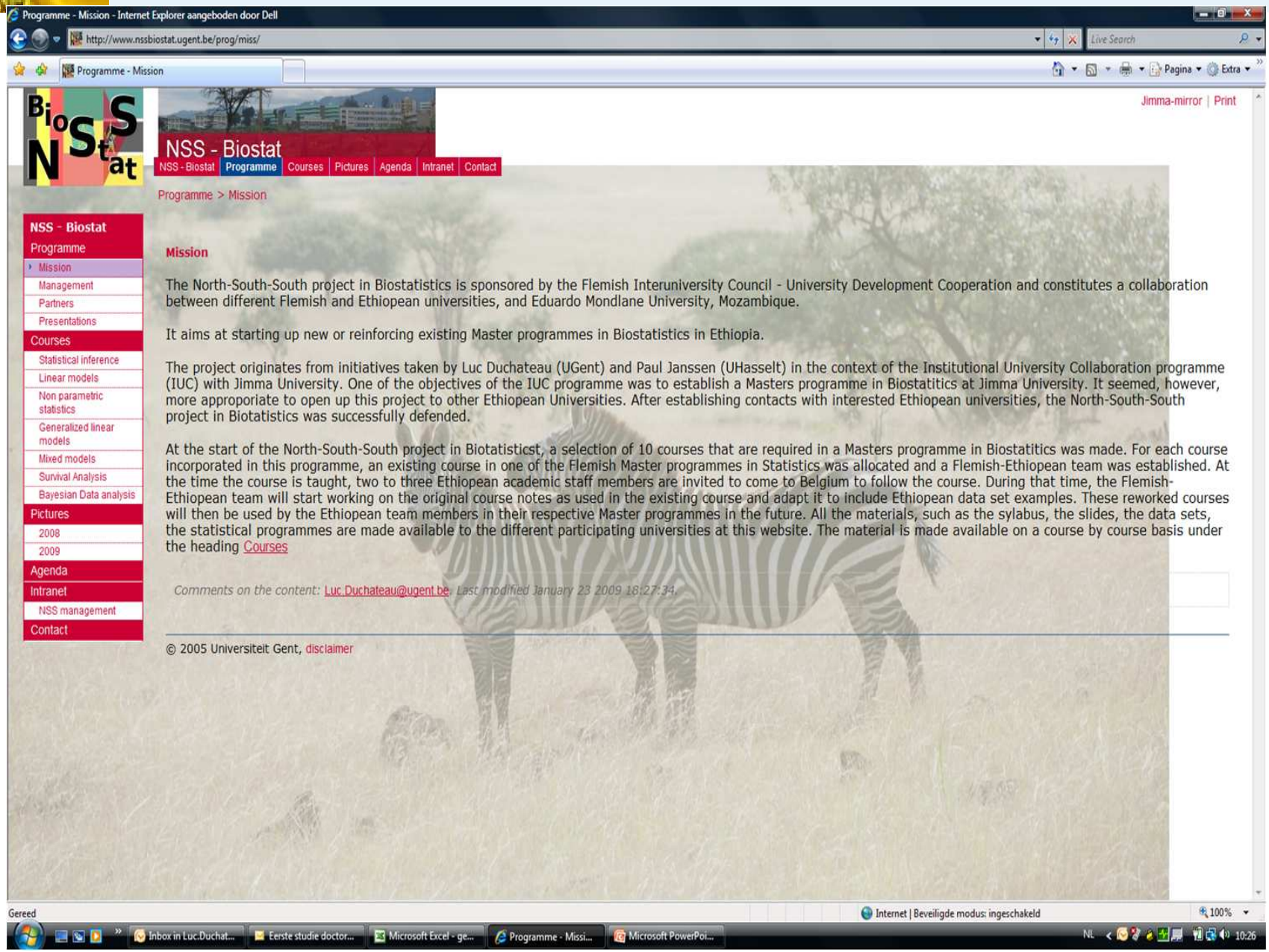
5.4. Workshops

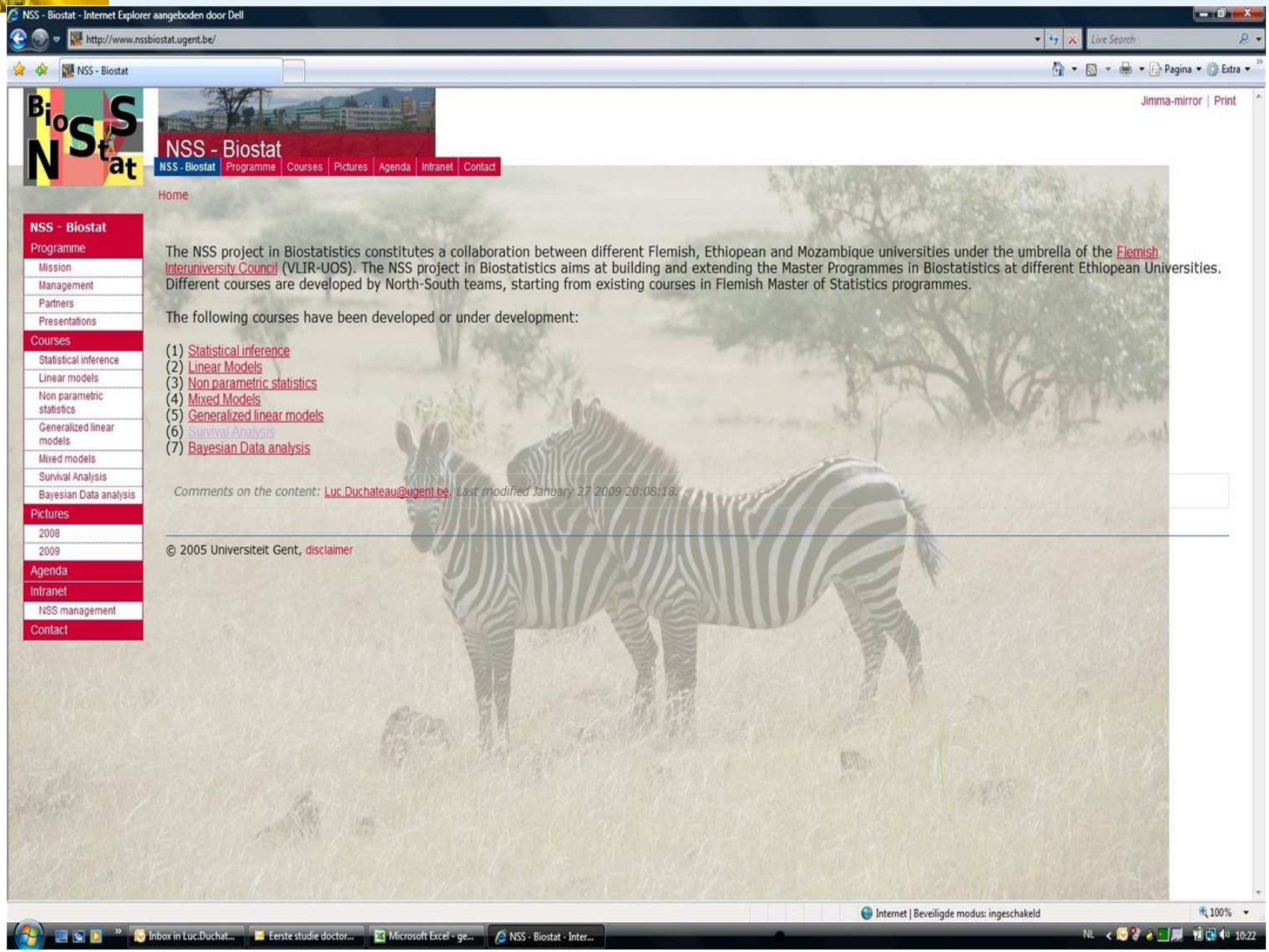
- La Molina workshop
  - Preparation phase, get to know one another
  - Distribution of tasks
  - Enlisting of experimental design problems
  - Drafting the structure for the coursenotes
  - Giving morning seminars
- Cuenca workshop (April 2012?)
  - Finalising list of experimental design problems
  - Exchanging the slides/coursenotes and working together towards finalisation
  - Trying out some of the materials in the coursenotes

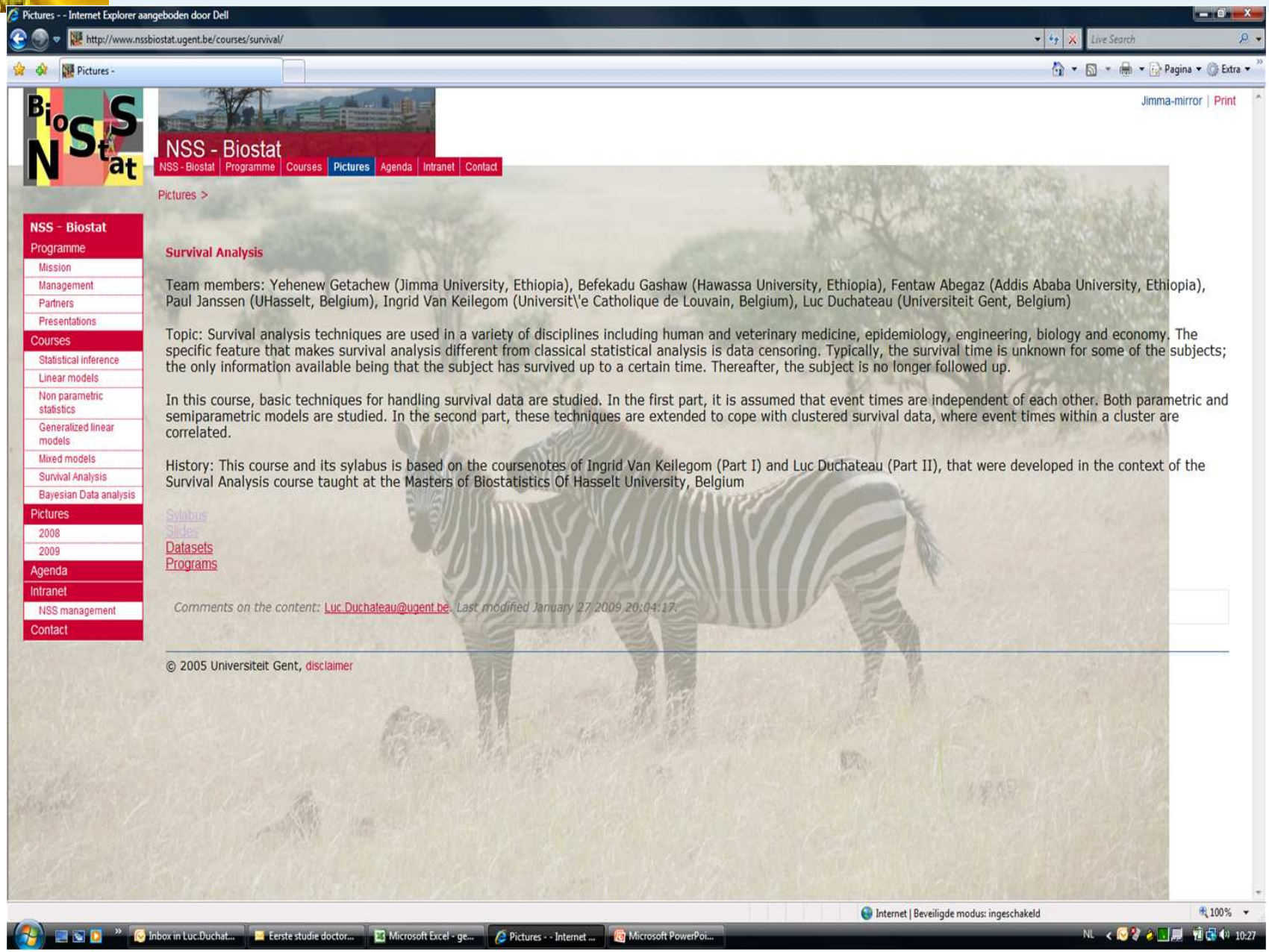
- Jimma workshop (February 2013?)
  - Giving the full courses to a large audience, inviting practitioners from different universities in Ethiopia, and few people from Cuenca and Lima (if budget allows).

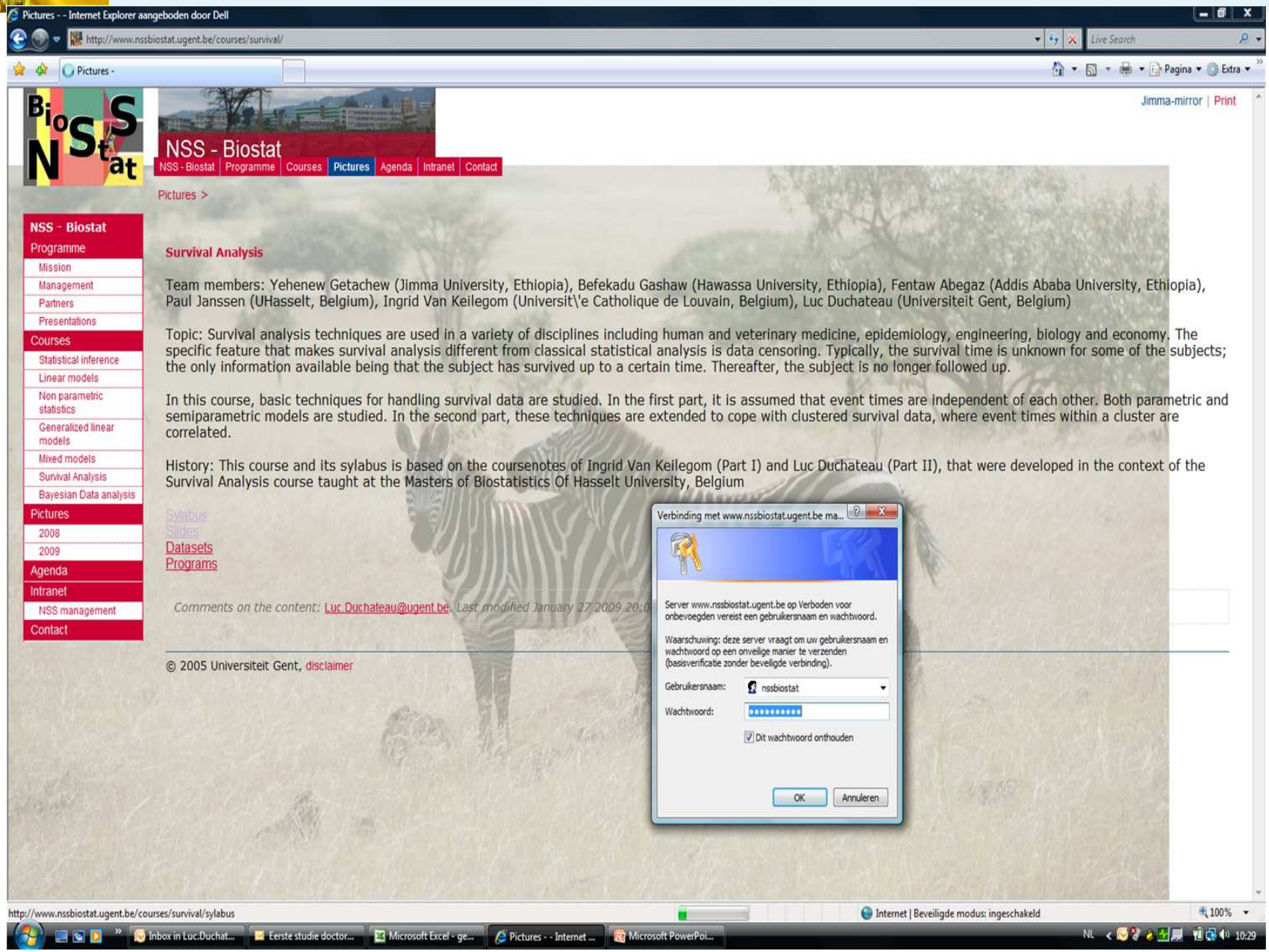
6. Predecessor: NSS biostatistics

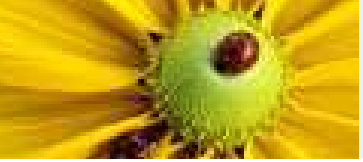



Programme - Mission - Internet Explorer aangeboden door Dell
http://www.nssbiostat.ugent.be/prog/miss/
Programme - Mission
Jimma-mirror | Print
NSS - Biostat
NSS - Biostat | Programme | Courses | Pictures | Agenda | Intranet | Contact
Programme > Mission
NSS - Biostat
Programme
Mission
Management
Partners
Presentations
Courses
Statistical inference
Linear models
Non parametric statistics
Generalized linear models
Mixed models
Survival Analysis
Bayesian Data analysis
Pictures
2008
2009
Agenda
Intranet
NSS management
Contact
Mission
The North-South-South project in Biostatistics is sponsored by the Flemish Interuniversity Council - University Development Cooperation and constitutes a collaboration between different Flemish and Ethiopean universities, and Eduardo Mondlane University, Mozambique.
It aims at starting up new or reinforcing existing Master programmes in Biostatistics in Ethiopia.
The project originates from initiatives taken by Luc Duchateau (UGent) and Paul Janssen (UHasselt) in the context of the Institutional University Collaboration programme (IUC) with Jimma University. One of the objectives of the IUC programme was to establish a Masters programme in Biostatitics at Jimma University. It seemed, however, more approporiate to open up this project to other Ethiopean Universities. After establishing contacts with interested Ethiopean universities, the North-South-South project in Biotatistics was successfully defended.
At the start of the North-South-South project in Biotatisticst, a selection of 10 courses that are required in a Masters programme in Biostatitics was made. For each course incorporated in this programme, an existing course in one of the Flemish Master programmes in Statistics was allocated and a Flemish-Ethiopean team was established. At the time the course is taught, two to three Ethiopean academic staff members are invited to come to Belgium to follow the course. During that time, the Flemish-Ethiopean team will start working on the original course notes as used in the existing course and adapt it to include Ethiopean data set examples. These reworked courses will then be used by the Ethiopean team members in their respective Master programmes in the future. All the materials, such as the sylabus, the slides, the data sets, the statistical programmes are made available to the different participating universities at this website. The material is made available on a course by course basis under the heading Courses
Comments on the content: Luc.Duchateau@ugent.be, Last modified January 23 2009 18:27:34.
© 2005 Universiteit Gent, disclaimer
Gereed
Internet | Beveiligde modus: ingeschakeld
100%
Inbox in Luc.Duchat...
Eerste studie doctor...
Microsoft Excel - ge...
Programme - Missi...
Microsoft PowerPoi...
10:26


NSS - Biostat - Internet Explorer aangeboden door Dell
http://www.nssbiostat.ugent.be/
NSS - Biostat
Jimma-mirror | Print
NSS - Biostat
NSS - Biostat | Programme | Courses | Pictures | Agenda | Intranet | Contact
Home
NSS - Biostat
Programme
Mission
Management
Partners
Presentations
Courses
Statistical inference
Linear models
Non parametric statistics
Generalized linear models
Mixed models
Survival Analysis
Bayesian Data analysis
Pictures
2008
2009
Agenda
Intranet
NSS management
Contact
The NSS project in Biostatistics constitutes a collaboration between different Flemish, Ethiopean and Mozambique universities under the umbrella of the Flemish Interuniversity Council (VLIR-UOS). The NSS project in Biostatistics aims at building and extending the Master Programmes in Biostatistics at different Ethiopean Universities. Different courses are developed by North-South teams, starting from existing courses in Flemish Master of Statistics programmes.
The following courses have been developed or under development:
(1) Statistical inference
(2) Linear Models
(3) Non parametric statistics
(4) Mixed Models
(5) Generalized linear models
(6) Survival Analysis
(7) Bayesian Data analysis
Comments on the content: Luc.Duchateau@ugent.be, Last modified January 27 2009 20:08:18.
© 2005 Universiteit Gent, disclaimer
Internet | Beveiligde modus: ingeschakeld
100%


Pictures - - Internet Explorer aangeboden door Dell
http://www.nssbiostat.ugent.be/courses/survival/
Live Search
Pictures -
Pagina
Extra
Jimma-mirror | Print
NSS - Biostat
NSS - Biostat
Programme
Courses
Pictures
Agenda
Intranet
Contact
Pictures >
NSS - Biostat
Programme
Mission
Management
Partners
Presentations
Courses
Statistical inference
Linear models
Non parametric statistics
Generalized linear models
Mixed models
Survival Analysis
Bayesian Data analysis
Pictures
2008
2009
Agenda
Intranet
NSS management
Contact
Survival Analysis
Team members: Yehenew Getachew (Jimma University, Ethiopia), Befekadu Gashaw (Hawassa University, Ethiopia), Fentaw Abegaz (Addis Ababa University, Ethiopia), Paul Janssen (UHasselt, Belgium), Ingrid Van Keilegom (Universit\'e Catholique de Louvain, Belgium), Luc Duchateau (Universiteit Gent, Belgium)
Topic: Survival analysis techniques are used in a variety of disciplines including human and veterinary medicine, epidemiology, engineering, biology and economy. The specific feature that makes survival analysis different from classical statistical analysis is data censoring. Typically, the survival time is unknown for some of the subjects; the only information available being that the subject has survived up to a certain time. Thereafter, the subject is no longer followed up.
In this course, basic techniques for handling survival data are studied. In the first part, it is assumed that event times are independent of each other. Both parametric and semiparametric models are studied. In the second part, these techniques are extended to cope with clustered survival data, where event times within a cluster are correlated.
History: This course and its sylabus is based on the coursenotes of Ingrid Van Keilegom (Part I) and Luc Duchateau (Part II), that were developed in the context of the Survival Analysis course taught at the Masters of Biostatistics Of Hasselt University, Belgium
Sylabus
Slides
Datasets
Programs
Comments on the content: Luc.Duchateau@ugent.be. Last modified January 27 2009 20:04:17.
© 2005 Universiteit Gent, disclaimer
Internet | Beveiligde modus: ingeschakeld
100%
Inbox in Luc.Duchat...
Eerste studie doctor...
Microsoft Excel - ge...
Pictures - - Internet ...
Microsoft PowerPoi...
NL
10:27


Pictures - - Internet Explorer aangeboden door Dell
http://www.nssbiostat.ugent.be/courses/survival/
Live Search
Pictures -
Pagina
Extra
Jimma-mirror | Print
NSS - Biostat
NSS - Biostat
Programme
Courses
Pictures
Agenda
Intranet
Contact
Pictures >
NSS - Biostat
Programme
Mission
Management
Partners
Presentations
Courses
Statistical inference
Linear models
Non parametric statistics
Generalized linear models
Mixed models
Survival Analysis
Bayesian Data analysis
Pictures
2008
2009
Agenda
Intranet
NSS management
Contact
Survival Analysis
Team members: Yehenew Getachew (Jimma University, Ethiopia), Befekadu Gashaw (Hawassa University, Ethiopia), Fentaw Abegaz (Addis Ababa University, Ethiopia), Paul Janssen (UHasselt, Belgium), Ingrid Van Keilegom (Universit\'e Catholique de Louvain, Belgium), Luc Duchateau (Universiteit Gent, Belgium)
Topic: Survival analysis techniques are used in a variety of disciplines including human and veterinary medicine, epidemiology, engineering, biology and economy. The specific feature that makes survival analysis different from classical statistical analysis is data censoring. Typically, the survival time is unknown for some of the subjects; the only information available being that the subject has survived up to a certain time. Thereafter, the subject is no longer followed up.
In this course, basic techniques for handling survival data are studied. In the first part, it is assumed that event times are independent of each other. Both parametric and semiparametric models are studied. In the second part, these techniques are extended to cope with clustered survival data, where event times within a cluster are correlated.
History: This course and its sylabus is based on the coursenotes of Ingrid Van Keilegom (Part I) and Luc Duchateau (Part II), that were developed in the context of the Survival Analysis course taught at the Masters of Biostatistics Of Hasselt University, Belgium
Sylabus
Slides
Datasets
Programs
Comments on the content: Luc.Duchateau@ugent.be. Last modified January 27 2009 20:0
© 2005 Universiteit Gent, disclaimer
Verbinding met www.nssbiostat.ugent.be ma...
Server www.nssbiostat.ugent.be op Verboden voor onbevoegden vereist een gebruikersnaam en wachtwoord.
Waarschuwing: deze server vraagt om uw gebruikersnaam en wachtwoord op een onveilige manier te verzenden (basisverificatie zonder beveiligde verbinding).
Gebruikersnaam:
nssbiostat
Wachtwoord:
Dit wachtwoord onthouden
OK
Annuleren
http://www.nssbiostat.ugent.be/courses/survival/sylabus
Internet | Beveiligde modus: ingeschakeld
100%
Inbox in Luc.Duchat...
Eerste studie doctor...
Microsoft Excel - ge...
Pictures - - Internet ...
Microsoft PowerPoi...
NL
10:29

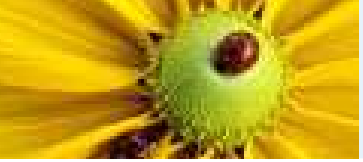


Survival Analysis

Yehenew Getachew
Befekadu Gashaw
Fentaw Abegaz
Paul Janssen
Ingrid Van Keilegom
Luc Duchateau

Academic Year 2009-2010
Master in Biostatistics. South West Ethiopia

This course has been developed within the context of the North-South-South project sponsored by VLIR-UOS. Belgium (see http:\\www.NSSbiostat.ugent.be)



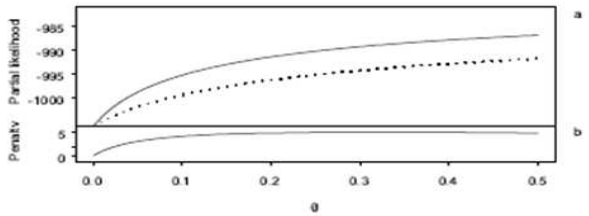
Yehenew Getachew
Befekadu Gashaw
Fentaw Abegaz
Paul Janssen
Ingrid Van Keilegom
Luc Duchateau

Survival Analysis

North-South-South project in Biostatistics Series

VLIR-UOS, Belgium

Figure 16.6: Profile penalised partial likelihood for θ: in (a) the penalised partial likelihood (dashed line) and the partial likelihood part (solid line); in (b) the penalty term.



$$\frac{\partial l_{pen}}{\partial w_a} = \frac{1 - \exp(w_a)}{\theta}$$

The partial derivative of the penalised partial likelihood with respect to w_a is

$$\frac{\partial l_{pal}}{\partial w_a} = d_a - H_{a,c}(\mathbf{v}_a)\exp(w_a) - \frac{1 - \exp(w_a)}{\theta} \qquad (16.16)$$

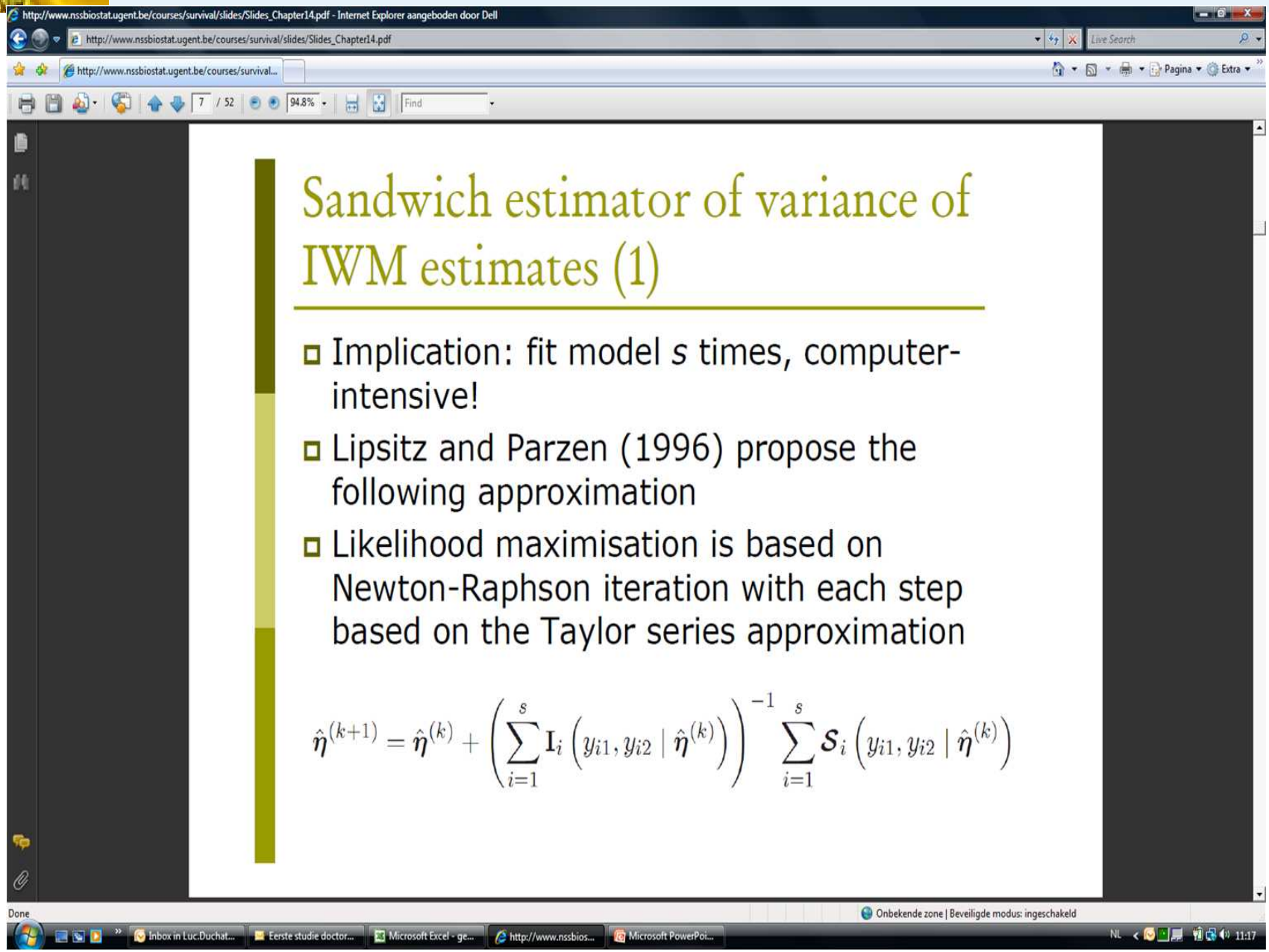
with

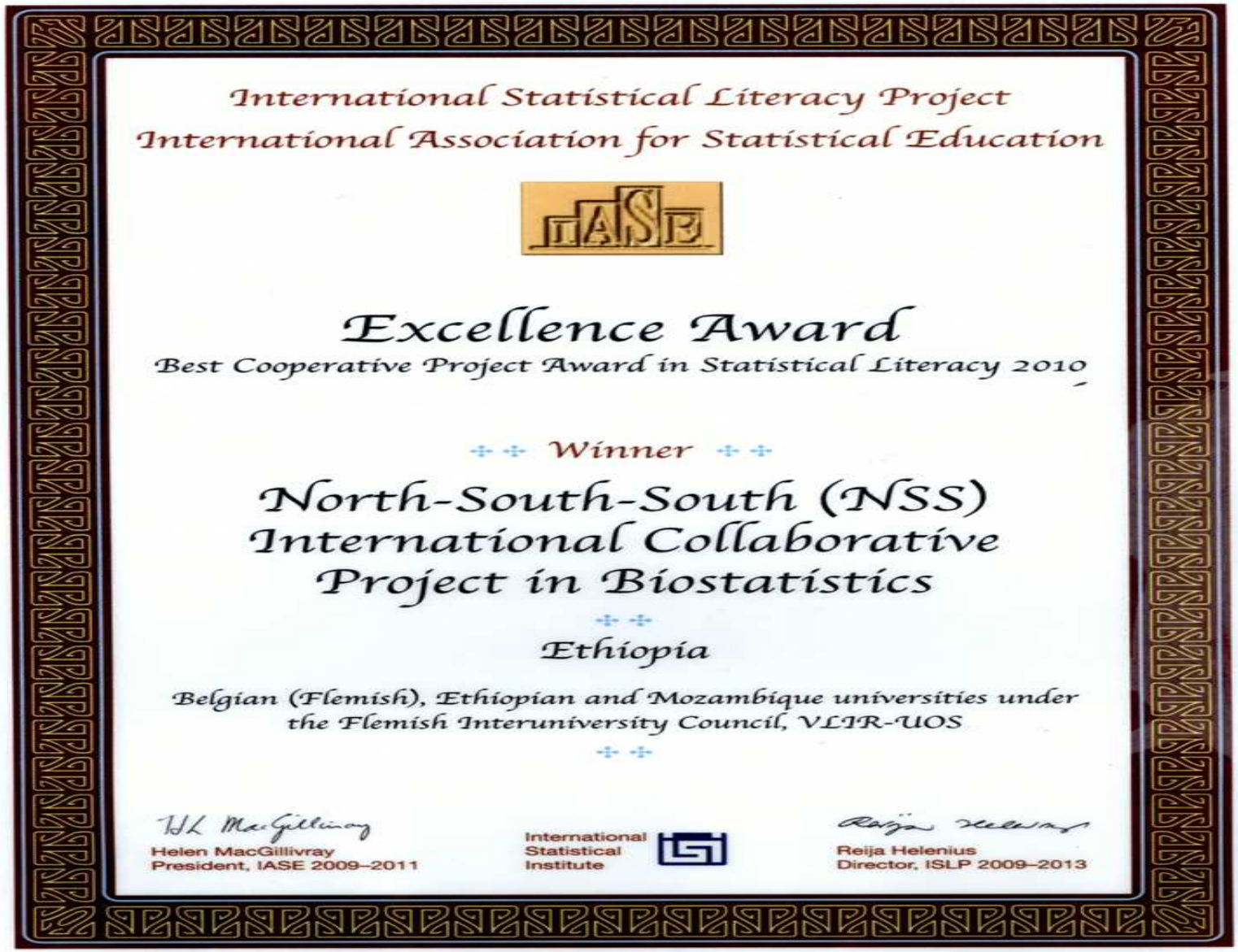
$$H_{a,c}(\mathbf{v}_a) = \sum_{i=1}^{n_a} \exp\left(\mathbf{x}_{ai}^t \boldsymbol{\beta}\right) H_0(u_{ai})$$

Now consider the solution for u_a from the modified EM algorithm for a particular value $\theta^{(l)}$. When the algorithm has converged at step $k+1$ within outer loop iteration l, the estimate for u_a, $\hat{u}_{a,\theta^{(l)}}$, can be taken as the expected value in (16.8), but with $\theta^{(k)}$ replaced by the fixed value for θ, $\theta^{(l)}$, as only the inner loop of the modified EM algorithm is currently considered. Similarly, we consider $H^{(k)}_{a,c}(\mathbf{v}_a)$ to be, upon convergence, the estimate $\hat{H}_{a,c,\theta^{(l)}}(\mathbf{v}_a)$. We need to convert the estimated frailty $\hat{u}_{a,\theta^{(l)}}$ to the random effect $\hat{w}_{a,\theta^{(l)}}$ to evaluate it in the context of the penalised partial likelihood approach, so denote by $\hat{w}_{a,\theta^{(l)}}$ the logarithm of the estimate $\hat{u}_{a,\theta^{(l)}}$.
We then have, by transforming (16.8), that

$$\hat{H}_{a,c,\theta^{(l)}}(\mathbf{v}_a) = \exp\left(-\hat{w}_{a,\theta^{(l)}}\right)\left(d_a + 1/\theta^{(l)}\right) - 1/\theta^{(l)} \qquad (16.17)$$

If we substitute this expression in the score equation (16.16) and evaluate it at the estimates $\boldsymbol{\beta}$ and $\mathbf{w}$ derived from the EM algorithm, we obtain


http://www.nssbiostat.ugent.be/courses/survival/slides/Slides_Chapter14.pdf - Internet Explorer aangeboden door Dell
http://www.nssbiostat.ugent.be/courses/survival/slides/Slides_Chapter14.pdf
Sandwich estimator of variance of IWM estimates (1)
Implication: fit model s times, computer-intensive!
Lipsitz and Parzen (1996) propose the following approximation
Likelihood maximisation is based on Newton-Raphson iteration with each step based on the Taylor series approximation
$\hat{\eta}^{(k+1)} = \hat{\eta}^{(k)} + \left(\sum_{i=1}^{s} \mathbf{I}_i \left(y_{i1}, y_{i2} \mid \hat{\eta}^{(k)} \right) \right)^{-1} \sum_{i=1}^{s} \mathbf{S}_i \left(y_{i1}, y_{i2} \mid \hat{\eta}^{(k)} \right)$
Done
Onbekende zone | Beveiligde modus: ingeschakeld



International Statistical Literacy Project
International Association for Statistical Education

IASE

Excellence Award
Best Cooperative Project Award in Statistical Literacy 2010

Winner

North-South-South (NSS)
International Collaborative
Project in Biostatistics

Ethiopia

Belgian (Flemish), Ethiopian and Mozambique universities under
the Flemish Interuniversity Council, VLIR-UOS

Helen MacGillivray
President, IASE 2009–2011

International
Statistical
Institute

Reija Helenius
Director, ISLP 2009–2013

7. Contemporary project: Cross cutting initiative in statistics

- Extends the output of NSS project in statistics towards Eastern and Southern Africa
- Eventual goals are
  - Further extension of cross cutting to all IUCs
  - Incorporation of NSS project in statistical research planning in cross cutting initiative

8. VLIR-UOS and statistics

- VLIR-UOS provides different statistical disciplines at two levels based on different collaborations