מפגש פורום ישראלי לביוסטטיסטיקה
Israeli Biostatistics Forum (IBF)

Tel-Aviv University, Tuesday, 26.11.13

שיטות לטיפול בנתונים חסרים
Methods for Dealing with Missing or Incomplete Data

9:00-9:30	***Registration and Coffee***
9:30-10:30	**Statistical methods for dealing with missing continuous data in longitudinal studies – application on data from a clinical trial on weight-loss interventions.** ***Michal Yackobovitch-Gavan, Tel-Aviv University***
10:30-11:30	**Model selection in incomplete data.** ***Ofer Harel, University of Connecticut, USA***
11:30-12:00	***Coffee break***
12:00-13:00	**Combining efficacy and completion rates with no data imputation: a composite approach with greater sensitivity for the statistical evaluation of active comparisons in antipsychotic trials.** ***Jonathan Rabinowitz, Bar Ilan University, Ramat-Gan***
13:00-14:00	***Lunch***
14:00-15:00	**Longitudinal and time-to-event joint modeling for handling cohort depletion in longitudinal aging study.** ***Diklah Geva, Ben Gurion University of the Negev***
15:00-16:00	**Viewing measurement error as a missing problem.** ***Larry (Laurence) Freedman, Gertner Institute, Tel-Hashomer***

דמי רישום סמליים יפורסמו בקרוב

הנחה לחברי EMR-IBS & ISA

לפרטים נוספים http://stat-athens.aueb.gr/~emribs/

גישות סטטיסטיות לטיפול בנתונים רציפים חסרים במחקרים אורכיים- יישום על נתוני מחקר קליני לטיפול בהשמנה

מיכל יעקובוביץ'-גוון (אוניברסיטת תל-אביב)
Email: michalyg2000@gmail.com

מבוסס על עבודת התיזה לקבלת התואר .MSc בביוסטטיסטיקה בהנחיית פרופ' דוד שטיינברג

תקציר:
ההרצאה תסקור בקצרה גישות לטיפול בנתונים רציפים חסרים במחקרים אורכיים, ביניהן שיטות השלמה שונות ומודלים שונים המבוססים על שיטת הנראות תחת הנחת MAR (Ignorable missing mechanism) ותחת הנחת MNAR (Non-ignorable missing mechanism).

בהמשך יודגם יישום של חלק משיטות אלה בניתוח מחקרה של פרופ' יעל בנימיני בנושא "השפעות מסוגלות עצמית ומניפולציה של הטמעת כוונות על ירידה במשקל בעקבות תוכנית טיפול קבוצתית" שנערך בשיתוף עם קופ"ח מכבי.

היישום של השיטות להתמודדות עם נתונים החסרים במחקר זה נעשה בשלושה מישורים:
1. בניית מודלים לאפיון הנשירה ובעזרתם קביעת מנגנון החסר (MNAR /MAR).

2. בניית מודלים להצלחת הטיפול (על פי המשתנה האורכי BMI) ובחירת המודל המתאים ביותר לנתונים על פי מבחני רגישות.

3. בחינת מספר המדידות המינימאלי ומיקומן, הדרושים לשם בניית מודלים אמינים לאפיון הצלחת הטיפול.

Model selection in incomplete data

Ofer Harel (University of Connecticut, USA)

Email: ofer.harel@uconn.edu

Abstract:

Model selection in complete data is a common task for the applied researcher. However, in many scenarios data are incomplete which further complicates the task of model selection. In this talk, we will specify the problem of model selection in incomplete data and discuss several possible solutions using multiple imputation.

First, we will define a new general measure for the correct model selection rates of common model selection criteria. Next, we will demonstrate the use of partial F-tests and define some new measures for model selection based on information criteria in multiply imputed data sets. This is a joint work with Ashok Chaurasia.

Keywords: Model Selection, Missing Data, Bayesian Analysis, Multiple imputation.

Combining efficacy and completion rates with no data imputation: a composite approach with greater sensitivity for the statistical evaluation of active comparisons in antipsychotic trials

Jonathan Rabinowitz, PhD_a, Nomi Werbeloff, PhD_a, Ivo Caers, PhD_b, Francine S. Mandel, PhD_c, Judith Jaeger, PhD_d, Virginia Stauffer, $PharmD_e$, François Menard, MD_f, Bruce J. Kinon, MD_e, Shitij Kapur, MD_g

Email: jrabin123@gmail.com

(a) Bar Ilan University, Ramat Gan, Israel; (b) Janssen Research & Development, Beerse, Belgium; (c) Pfizer; 235 East 42nd Street, New York, NY, 10017, USA; (d) Astra-Zenecca; (e) Eli Lilly, 893 South Delaware Street, Indianapolis, IN 46285, USA; (f) Lundbeck SAS; 37-45, Quai du Président Roosevelt, 92445, Paris, France; (g) Institute of Psychiatry, Kings College, London De Crespigny Park, London, UK, SE5 8AF

Acknowledgement: The research leading to these results has received support from the Innovative Medicine Initiative Joint Undertaking under grant agreement n° 115008 of which resources are composed of EFPIA in-kind contribution and financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013).

Abstract:

Background: Outcomes in RCT's of antipsychotic medications are often examined using last observation carried forward (LOCF) and mixed effect models (MMRM), these ignore meaning of non-completion and thus rely on questionable assumptions. We tested an approach that combines into a single statistic, the drug effect in those who *complete* trial and *proportion* of patients in each treatment group who complete trial. This approach offers a conceptually and clinically meaningful endpoint.

Objective: Composite approach was compared to LOCF (ANCOVA) and MMRM in 59 industry sponsored RCT's. *Methods:* For within study comparisons we computed effect size (z-score) and p values for (a) rates of completion, (b) symptom change for complete cases, which were combined into composite statistic, and (c) symptom change for all cases using last observation forward (LOCF). *Results*: In the 30 active comparator studies, composite approach detected larger differences in effect size than LOCF (ES=.05) and MMRM (ES=.076). In 10 of the 49 comparisons composite lead to significant differences (p<=.05) where LOCF and MMRM did not. In 3 comparisons LOCF was significant, in 2 MMRM lead to significant differences whereas composite did not.

In placebo controlled trials, there was no meaningful difference in effect size between composite and LOCF and MMRM when comparing placebo to active treatment, however composite detected greater differences that other approaches when comparing between active treatments.

Conclusions: Composite was more sensitive to effects of experimental treatment versus active controls (but not placebo) than LOCF and MMRM thereby increasing study power while answering a more relevant question.

Longitudinal and time-to-event joint modeling for handling cohort depletion in longitudinal aging study; HEALTH ABC

Diklah Geva[1*†], Danit Shahar[1], Tamara Harris[2], Geert Molenberghs[3], Michael Friger[1]

[1] Public Heath, Faculty of Health Science, Ben Gurion University of the Negev, P.O. Box 653 Beer-Sheva Israel.
[2] Laboratory of Epidemiology, Demography, and Biometry, 7201 Wisconsin Av., Bethesda, MD 20892 U.S.A.
[3] Center for Statistics (CenStat), Universiteit Hasselt, Agoralaan 1, B-3590 Diepenbeek, Belgium.

E-mail*:* gevadi@post.bgu.ac.il

Abstract:

Background: Longitudinal cohort studies focused on older people are prone to selection bias due to the increased attrition as follow-up continues. This is an important issue that requires analytical attention. In the past decades, three schools have evolved to handle the "missing data" problem: semi-parametric approaches, multiple imputations, and joint modeling. The 3 main branches of the joint modeling school are selection models, pattern-mixture models, and shared-parameter models. This presentation focuses on the last family, which accommodate continuous dropout time.

Objective:

1) To present the different joint modeling approaches and to discuss availability in R.
2) To present a joint modeling application to analysis of the association between muscle strength and gait speed over 10 years from data of the Health ABC cohort study.

Application to data: Initial data analysis includes graphical mapping of missing data and generation of a missing data index (MDI). Next, the main predictor was classified according to its trend over time using a heterogynous latent class mixed model. The main analysis is the joint modeling of two processes: walking speed (m/sec) and time to death/censoring. This was followed by a sensitivity analysis including spline fit, Lag fit, MDI stratification and impact of varying time windows for latent class generation. Additional validation was obtained using simulation studies.
The R-code used for generating analysis will be presented and includes the lcmm and JM packages.

Conclusions: Despite the recognition that missing data is a major problem in studies of older people, to our knowledge this is the first attempt to use a joint modeling approach to study walking speed in elderly. We have illustrated significant association between the longitudinal and time to event process using joint models and studied the impact of it on walking speed muscle strength relationship in the Health ABC cohort.

Key words: Missing Data, Geriatric Cohort Studies, Joint Models, Longitudinal Analysis

Viewing measurement error as a missing data problem

Laurence S. Freedman (Gertner Institute)

E-mail: lsf@actcom.co.il

Abstract:

Problems involving the use of measurements that are imprecisely reported or measured have been discussed extensively in the statistical and biostatistical literature, and many different methods have been proposed. The main type of problem that has been addressed, is where we are interested in the regression relationship of an outcome variable Y on an explanatory variable X, but where we cannot observe X (except possibly in a relatively small subset of individuals) and instead observe an imperfect measure X, which is usually denoted W. In this context, the following methods, among others, have been proposed: full maximum likelihood, conditional scores, SIMEX, regression calibration (RC), moment reconstruction (MR) and multiple imputation (MI). Of these, the latter three methods can all be viewed as imputation methods in which the unknown X for each individual is imputed via knowledge of W. In such situations, the variable X is missing in 100% or close to 100% of the individuals, and nevertheless RC, MR and MI are all acceptable methods of dealing with the missingness!

In RC, X is substituted by the conditional expectation E(X|W), based on the assumption that the error in the measurement of X is non-differential, i.e. independent of Y conditional on X. In moment reconstruction (MR), X is imputed using a variable that is constructed to equal the first and second moments of X and its covariance with Y. In MI, X is imputed through a model relating X to W and Y. Neither MR nor MI rely on the non-differential error assumption. We investigated these methods using computer simulations. Our results show that for most situations in epidemiology, RC is preferable when there is non-differential measurement error. Under this condition, there are cases where RC is less efficient than MR or IM, but they rarely occur in epidemiology. We show that the efficiency gain of usual RC over the other methods can sometimes be dramatic. When differential measurement error does pertain, then MR and IM have considerably less bias than RC, but can have much larger variance. Versions of MR and IM can be derived that use the non-differential error assumption, and these versions perform similarly to RC. We demonstrate our findings with an analysis of dietary fat intake and mortality in a large cohort study.

References

1. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models: A Modern Perspective.* 2nd edition. Chapman Hall/CRC, Boca Raton, FL, 2006.

2. Freedman LS, Fainberg V, Kipnis V, Midthune D, Carroll RJ. A new method for dealing with measurement error in explanatory variables of regression models. *Biometrics* 2004; **60**:172-181.

3. Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. *International Journal of Epidemiology* 2006; **35**:1074-1081.

4. Freedman LS, Midthune D, Carroll RJ, Kipnis V. A comparison of regression calibration, moment reconstruction and imputation for adjusting for covariate measurement error in regression. *Statistics in Medicine* 2008; 27:5195-5216.