Nieuwsbrief no. 41
November 2008

Vereniging voor Ordinatie en Classificatie

Voorzitter: Ron Wehrens, Radboud University Nijmegen, Analytical Chemistry, Toernooiveld 1, 6525 ED Nijmegen (r.wehrens@science.ru.nl)
Secretaris: Hugo Duivenvoorden, Erasmus Universiteit, Fac. Geneeskunde en Gezondheidswetenschappen, Postbus 1738, 3000 DR Rotterdam (h.duivenvoorden@erasmusmc.nl)
Penningmeester: Laurence Frank, Disciplinegroep M&T, Fac. Soc. Wet., Universiteit Utrecht, Postbus 80140, 3508 TC Utrecht (L.E.Frank@fss.uu.nl). Postbankrekening 161723 t.n.v. Vereniging voor Ordinatie en Classificatie, Naarden. Bankrekening nummer 777-5952385-56 Dexia Bank t.n.v. VOC, Naarden
Redactie: Eva Ceulemans, Centrum voor methodologie van het pedagogisch onderzoek, Katholieke Universiteit Leuven, Vesaliusstraat 2, B-3000 Leuven, België (Eva.Ceulemans@ped.kuleuven.be)
VOC-home page: http://www.voc.ac

Fall Meeting of the VOC and Werkgroep Medische Statistiek

November 28, 2008

Erasmus Universiteit Rotterdam, Campus Woudestein, T-Gebouw (T3-24)

10.00	Registration and Coffee
10.30	Saskia le Cessie: *Propensity scores, an introduction*
11.30	Fannie Cobben: *Using response propensities in survey research*
12.15	Lunch
13.45	Arjan Blokland: *The (collateral) effects of imprisonment*
14.30	Edwin Martens: *Preference for propensity scores when estimating an average treatment effect in case of a dichotomous outcome*
15.15	Tea
15.45	Stef van Buuren: *Pooling outcomes after quintile stratification*
16.30	Drinks

In this issue:

Registration details for the Fall Meeting:

Those who would like to participate are welcome and are kindly requested to register via the VOC website (http://www.voc.ac). Participation is free. Registration deadline: November the 24th.

From the President

This August, the German Classification Society (GfKl) organised its 32nd yearly conference in Hamburg, which was attended by nearly 300 participants. The VOC and the British Classification Society were co-organisers, both hosting a special session. In the VOC session, Andries van der Ark, Peter van der Heijden and Marieke Timmermans gave interesting talks. Next year's GfKl meeting will coincide with the IFCS conference, held from March 13--18 in Dresden, again an outstanding opportunity to listen and talk to many scientists in the field of classification.

Before that, however, there are other meetings worth mentioning, not in the least our own fall meeting. This will be held in Rotterdam on November 28 and will present an interesting series of talks about propensity scores. You can find a complete programme and abstracts elsewhere in this newsletter. I hope to see many of you there!

Ron Wehrens

IFCS 2009: Chikio Hayashi Awards Program

Procedure for application and nomination

The IFCS (International Federation of Classification Societies) provides awards to single persons in the framework of its Chikio Hayashi Awards Program (CHA). Awards are given to promising researchers who will present a paper on classification, data analysis and related areas at the IFCS 2009 conference, and who are in the early stages of their professional careers, as a support for attending an IFCS conference. Candidates shall be under the age of 35 years at the time of the conference. The award is a cash payment to a candidate who attends the IFCS conference and may be used for travel, accommodation, and conference costs.

Timetable

November 30, 2008:

Candidates must have submitted their CHA application to the chair of the Awards Committee (Anuska Ferligoj). The application form can be downloaded from the IFCS 2009 website.

January 20, 2009:

The Awards Committee proposes prize winners to the IFCS President for approval, and if the President approves, the prize winners will be informed by the chair of the Awards Committee (Anuska Ferligoj). Full details are given in the CHA Guidelines that can be downloaded from the IFCS 2009 website.

Abstracts for the Fall Meeting

Saskia le Cessie (Department of Medical Statistics and Department of Clinical Epidemiology, LUMC, Leiden): Propensity scores, an introduction

Using propensity scores to deal with confounding has become very popular in recent years. By estimating the probability to receive a certain treatment (the propensity), one can adjust for observed imbalance between treatment groups. In this talk, the basis concepts of propensity scores are considered. We discuss in which situations propensity scores are useful. We also consider how propensity scores can be constructed; answering questions like whether all possible variables related to the treatment should be included in the score. Finally we compare different ways of using propensity scores: propensity matching, stratification, inverse probability weighting, and using the propensity score as covariate. We show that the different approaches can yield quite different results.

Saskia le Cessie *is an associate professor at the department of Medical Statistics and Bio-informatics and at the department of Clinical Epidemiology of the Leiden University Medical Center. Her research interests are in statistical methods for epidemiological research. She is a consultant for the Comprehensive Cancer Center West and a member of the scientific board of the Dutch Arthritis Association (Reumafonds). She has been an associate editor of Applied Statistics (JRSS-C) and has served in the Editorial Advisory Committee and the Council of the International Biometrical Society. She has been a co-author of over 150 publications in Medical and Statistical Journals.*

Fannie Cobben (Statistics Netherlands, Division of Methodology and Quality): Using response propensities in survey research

The idea of using response probabilities in survey research has received much attention in survey methodological literature lately due to the introduction of the propensity score method. The propensity score method originates from evaluation studies to estimate average treatment effects and was first introduced by Rosenbaum and Rubin (1983). In treatment effect studies, there usually are two groups involved: one group that receives the treatment, and one group that serves as a control group and does not receive the treatment. The statistic of interest is the effect of the treatment. However, to unbiasedly measure this effect, it is necessary to remove all possible differences in outcome that arise due to a different composition of the treatment and the control group. For this purpose, the propensity score is introduced as a way to balance the composition of the two groups.

The idea of using the propensity score method in survey methodology was first introduced by Harris Interactive; see Taylor et al. (2001). Harris Interactive uses the propensity score method to solve problems of under-coverage and self-selection in volunteer opt-in internet

panels. In this perspective, the propensity score method relies on the availability of a reference survey that does not have the same problems of under-coverage and self-selection. This reference survey is used as a benchmark for the internet respondents by balancing the composition of the attributes of the web respondents so that they are similar to the composition of the attributes of the reference survey respondents. The propensity score is determined as the conditional probability that a sample element responds to the internet panel, given the attributes and the fact that the element responded either in the reference survey or the internet panel.

The estimated response probability can also be regarded as a propensity score. The use of (inverse) response probabilities to adjust for nonresponse bias has been introduced by Horvitz and Thompson already in 1952. They suggest adjusting the inclusion probabilities in the Horvitz-Thompson estimator for (selective) nonresponse. Amongst others, Bethlehem (1988) and Särndal et al. (1992) describe how the Horvitz-Thompson estimator can be modified to fit the situation of survey nonresponse by using estimated response probabilities.

In my presentation I describe methods that use response propensities to adjust for nonresponse bias. I will describe the applicaton of these methods to the Integrated Survey of Household Living Conditions 2002. Furthermore, I will discuss different aspects of the methods, and compare these methods to the traditional nonresponse adjustment technique of linear weighting.

References

Bethlehem, J.G. (1988). Reduction of Nonresponse Bias through Regression Estimation. *Journal of Official Statistics, 4*, 251-260.

Horvitz, D.G., & Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association, 47,* 663-685.

Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrica, 70*, 41-55.

Särndal, C.E., Swensson, B., & Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer-Verlag, New York.

Taylor, H., Bremer, J., Overmeyer, C., Siegel, J.W., & Terhanian, G. (2001). The Record of Internet-Based Opinion Polls in Predicting the Results of 72 Races in the November 2000 U.S. Elections. *International Journal of Market Research, 43,* 127 – 136.

Fannie Cobben *studied econometrics at the University of Amsterdam (VU). After obtaining her masters degree, she started working in 2004 at Statistics Netherlands, Division of Methodology and Quality. Her PhD-research, which is supervised by Jelke Bethlehem, focuses on the analysis and correction of nonresponse in survey research.*

Arjan Blokland (Netherlands Institute for the Study of Crime and Law Enforcement, Leiden): The (collateral) effects of imprisonment

At the outset of the new millennium 2.5 million individuals are confined in prisons or jails across North America and Western Europe and in most countries rates are at or near all time highs. A growing international literature has been attentive to the collateral consequences of the increased use of imprisonment. The potential irony of mass imprisonment is that, to the extent it has unintended adverse effects on life outcomes that are correlated with criminal offending, large-scale growth in the incarceration rate may actually exacerbate the crime problem over the long run by stigmatizing an ever larger class of individuals. Using data from the Netherlands-Based Criminal Career and Life-course Study the effect of first-time adult imprisonment on criminal recidivism and life circumstances in the years following the imprisonment was examined. Unadjusted comparisons of those imprisoned and those not imprisoned will be biased because imprisonment is not meted out randomly. Selection processes will tend to make the imprisoned group disproportionately crime prone compared to the not imprisoned group. In this study group-based trajectory modeling was combined with risk set matching to balance a variety of measurable indicators of criminal propensity.

Arjan Blokland *(PhD) is researcher at the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR) in Leiden and senior-researcher at Parnassia Addiction Research Centre (PARC) in The Hague. In 2006 he received a VENI-grant for his work on specialization in offending. He currently chairs the European Developmental and Life-course Criminology working group. His main area of research is life course criminology and focuses on the development of criminal careers, the influences of life course transitions on criminal behavior, drug use and crime and the (un)intended consequences of interventions.*

Edwin Martens (Utrecht University): Preference for propensity scores when estimating an average treatment effect in case of a dichotomous outcome

In observational studies with a dichotmous outcome, a multivariable logistic regression analysis is often used to adjust for confounding and estimate an adjusted treatment effect. This treatment effect is in general an overestimation of the treatment effect that is in most circumstances the intended one. The method of propensity scores on the other hand, will result in a treatment effect that is in general closer to the treatment effect that would have been found when the study was a randomized one. The larger the number of confounders or the larger the treatment effect, the more preferred is the method of propensity scores over a multivariable logistic regression analysis.

After 12 years of sociologic-economic research at the Erasmus University of Rotterdam, ***Edwin Martens*** *worked from 2000 as a biostatistician at Utrecht University. In 2007 he finished his PhD on the methods of propensity scores and instrumental variables.*

Stef van Buuren (TNO Quality of Life, Leiden & Department of Methodology and Statistics, FSS, University of Utrecht): Pooling outcomes after quintile stratification

Propensity score methods offer both theoretical and practical advantages over conventional regression techniques to control for bias in observational studies. Quintile stratification is a popular technique in which exposed and non-exposed subjects are divided into five homogeneous strata. Exposed and non-exposed are compared within each stratum, which leads to five results instead of one. The relevant literature pays surprising little attention to the problem how to aggregate these results into one overall estimate. I will outline pooling methods for differences in means and proportions and for the odds ratio, and illustrate these methods on real data.

Stef van Buuren *develops and applies quantitative methods in medicine and social science, with an emphasis on childhood growth and incomplete data. Van Buuren is professor of applied statistics in prevention at the University of Utrecht. More information can be found at http:\\www.stefvanbuuren.nl.*

Book review

Applied Multiway Data Analysis. Pieter M. Kroonenberg. Hoboken (NJ): John Wiley & Sons, Inc. 2008.

The term 'Multiway analysis' refers to component modeling of multiway data – data that can be organized in multidimensional boxes. For example, threeway data result when a number of individuals are repeatedly measured on a number of variables in a number of conditions. To have multiway data it is essential that the design is fully crossed. The usually complex relationships between individuals, variables and conditions can be disentangled with some kind of multiway component analysis. Such an analysis is not straightforward, because quite some nontrivial decisions have to be made, like the type of preprocessing of the observed data before analysis, and the specific model to use. The interpretation of the analysis results is usually not straightforward either, which is intrinsic to the complex data structures involved. However, commonly the researcher is rewarded with insights into the data that could not be obtained with other methods. The relatively unfamiliarity of multiway models is an extra barrier to use multiway analysis in practice. This book is an important step in making multiway analysis better accessible to a wider audience than the multiway research community itself.

As stated in the Preface 'The core of this book is concerned with practical issues in applying multiway component techniques…'. Although this book is clearly guided by the practical issues, theory is certainly not avoided. This is a fortunate choice. A profound theoretical background is necessary to obtain a good understanding of the issues involved, and hence to make proper choices in performing a multiway analysis. The theory is carefully introduced, both at the mathematical and conceptual levels. Considerable effort has been put to make the theory understandable: The matter is often explained from different viewpoints, and supplemented with insightful graphs. I like this very much: it helps novices to obtain a good understanding; to readers who are familiar with multiway analysis it may shed a different light, as it did to me at certain places. When technical details are being short cut too much to the taste of the reader, as it was for me at some places, (s)he can resort to the original literature. The references are numerous and appropriate. As such, the book provides a very good overview of the available literature on threeway analysis, both on the theoretical and applied sides. The main application area of this book is the social sciences, which contrasts to the other book on multiway analysis, *Multi-way Analysis with Applications in the Chemical Sciences* (2004, by Smilde, Bro and Geladi). Furthermore, the latter is somewhat more technically oriented. Hence, both books nicely complement each other.

For absolute novices, the book may be somewhat discouraging. The general ins and outs of multiway analyses are discussed in Part I, on Data, Models, and Algorithms, and Part II, on Data handling, Model selection, and Interpretation, which take as much as 308 pages. This is quite some investment to get a feeling for multiway analysis. Furthermore, readers totally unfamiliar with multiway analysis may find it difficult to see the outline and to select the main points. Both problems could have been prevented by starting with an overview example, with references to coming chapters. As an alternative, introductory articles can partly serve this purpose, like the one of Kiers and Van Mechelen (2001, in *Psychological Methods*). Another problem for absolute novices may be the order of presentation. In particular, in the empirical examples in Chapter 7 some novel terms are used without a –short– explanation, or reference to a future section. This could have done rather easily, and is preferable over the alternative – adopting an Agatha Christie's style of ordering, and postponing all empirical examples towards the end of the book.

Good things take time – and this holds for this book as well. The idea for the present book dates back to 1996. In my view, it was worth that time. Pieter Kroonenberg is one of the few with a profound knowledge of multiway analysis. It is meritorious that he took the effort to share his knowledge. It is to be hoped that a next edition will appear soon, to resolve some

typo's and inaccuracies in the current edition, but above all because the book deserves a broad reading public.

Marieke Timmerman, University of Groningen, The Netherlands

Personalia

Marike Polak started working as assistant professor at the Erasmus Universiteit Rotterdam, M&T, Psychologie, e-mail: polak@fsw.eur.nl

Agenda

The theme of the fall meeting of the Social Sciences Division of the Netherlands Statistical Society is Web Surveys. The meeting takes place on November 26nd in Enschede at the University of Twente. More information can be found at www.vvs-or.nl/sws/sws.html

On December 10 en 11, Dr. Joerg Henseler & Dr. Christian M. Ringle organize a seminar on Partial Least Square (PLS) Path Modeling - Using SmartPLS. Location of the seminar is Telekom Tagungshotel Neuss (near Düsseldorf), Germany. More information can be found here: http://www.pls-school.com

January 26 – 28. **Houston, Texas, USA.** 2009 Bayesian Biostatistics. www.mdanderson.org/departments/biostats/

March 13 - 18. **Dresden, Germany.** IFCS 2009: 11th Conference of the International Federation of Classification Societies. http://www.ifcs2009.de/

March 18 – 20. **Hong Kong, Hong Kong**. IAENG International Conference on Data Mining and Applications 2009. www.iaeng.org/IMECS2009/ICDMA2009.html

March 24 – 27. **Tokyo, Japan.** The 6th International Conference on Multiple comparison Procedures. www.mcp-coference.org

April 16 – 18. **Clearwater, Florida, USA.** Twelfth International Conference on Artificial Intelligence and Statistics. www.ics.uci.edu/~aistats/

April 30 – May 2. **Reno, Nevada, USA.** SDM09: The Ninth SIAM International Conference on Data Mining. www.siam.org/meetings/sdm09

May 3 – 8. **Ascona, Switzerland.** Statistical Advances in Genome-scale Data Analysis. www.stat.ethz.ch/talks/Ascona_09

May 25 – 29. **Bordeaux, France.** 41st annual conference of the French Statistical Society. www.sm.u-bordeaux2.fr/JDS2009/index.html

May 27 – 29. **Vancouver, British Columbia, Canada.** 2009 Spring Research Conference on Statistics in Industry and Technology. www.stat.sfu.ca/~boxint/src2009/

June 2 – 5. **Yorktown Heights, New York, USA.** 2009 Quality & Productivity Research Conference. www.research.ihost.com/qprc_2009/

June 5 – 9. **Philadelphia, Pennsylvania, USA.** O-Bayes09. www.stat.wharton.upenn.edu/statweb/Conference/OBayes09/OBayes.html

June 18 – 20. **Bressanone/Brixen, Italy.** BISP6 - Bayesian Inference in Stochastic Processes. www.mi.imati.cnr.it/conferences/bisp6.html

June 21 – 25. **San Francisco, California, USA.** ICSA 2009 Applied Statistics Symposium. www.icsa2.org/2009

June 21 – 25. **Moncalieri (Turin), Italy.** 7th Workshop on Bayesian Nonparametrics. www.bnpworkshop.carloalberto.org

June 23 – 27. **Smolenice Castle, Slovakia.** 18th International Workshop on Matrices and Statistics, IWMS'09. http://www.um.sav.sk/en/iwms2009.html

June 25 – 27. **Columbus, Ohio, USA.** United States Conference on Teaching Statistics (USCOTS) '09. www.causeweb.org/uscots

July 1 – 3. **London, UK.** International Conference of Computational Statistics and Data Engineering 2009. www.iaeng.org/WCE2009/ICCSDE2009.html

July 20 – 22. **St. John's, Newfoundland, Canada.** International Symposium in Statistics on GLLMM. www.iss-2009-stjohns.ca

July 20 – 24. **Cambridge, UK.** IMPS 2009: International meeting of the Psychometric Society. http://www.psychometrika.org/meeting/2009/leaflet.pdf

August 1 – 6. **Washington, District of Columbia, USA.** 2009 Joint Statistical Meetings. www.amstat.org/meetings/

August 16 – 22. **Durban, South Africa.** International Statistical Institute 57th Biennial Session in Durban. www.cbs.nl/isi/

Publications

Bennani-Dosse, M., & Ten Berge, J.M.F. (2008). The assumption of proportional components when Candecomp is applied to symmetric matrices in the context of Indscal. *Psychometrika, 73*, 303-307.

Bijleveld, C.C.J.H., & Commandeur, J.J.F. (2008). *Multivariate analyse. Een inleiding voor criminologen en andere sociale wetenschappers.* Den Haag: Boom Juridische uitgevers.

Ceulemans, E., & Van Mechelen, I. (2008). CLASSI: A classification model for the study of sequential processes and individual differences therein. *Psychometrika, 73*, 107-124.

de Jong, K., Nugter, M.A., Polak, M.G., Wagenborg, J.E.A., Spinhoven, P., & Heiser, W.J. (2008). De Nederlandse versie van de Outcome Questionnaire (OQ-45): een cross-culturele validatie. *Psychologie & Gezondheid, 36*, 35-45.

del Pino, G., San Martín, E., González, J., & De Boeck, P. (2008). On the relationships between sum score based estimation and joint maximum likelihood estimation. *Psychometrika, 73*, 145-151.

Depril, D., Van Mechelen, I., & Mirkin, B.G. (2008). Algorithms for additive clustering of rectangular data tables. *Computational Statistics & Data Analysis, 52*, 4923-4938.

de Vries, A.C., van Grieken, N.C., Looman, C.W.N., Casparie, M.K., de Vries, E., Meijer, G.A., & Kuipers, E.J. (2008). Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. *Gastroenterology, 134*, 945-52.

Frank, L.E., & Heiser, W.J. (2008). Feature selection in feature network models: Finding predictive subsets of features with the positive lasso. *British Journal of Mathematical and Statistical Psychology, 61,* 1-27.

Goegebeur, Y., De Boeck, P., Wollack, J.A., & Cohen, A.S. (2008). A speeded item response model with gradual process change. *Psychometrika, 73*, 65-87.

Groenen, P.J.F., Nalbantov, G., & Bioch, J.C. (2008). SVM-Maj: a majorization approach to linear support vector machines with different hinge errors. *Advances in Data Analysis and Classification, 2*, 17-43.

Hickendorff, M., Heiser, W.J., Van Putten, C.M., & Verhelst, N.D. (2008). Clustering nominal data with equivalent categories. *Behaviormetrika, 35*, 35-54.

Leenen, I., Van Mechelen, I., Gelman, A., & De Knop, S. (2008). Bayesian hierarchical classes analysis. *Psychometrika, 73*, 39-64.

Nusselder, W.J., Looman, C.W.N., Franco, O.H., Peeters, A., Slingerland, A.S., & Mackenbach, J.P. (2008). The relationship between physical activity and years lived with and without disability. *Journal of Epidemiology and Community Health, 62*, 823-828.

Polek, E., Van Oudenhoven, J.P., & Ten Berge, J.M.F. (2008). Attachment styles and demographic factors as predictors of sociocultural and psychological adjustment of Eastern European emigrants in the Netherlands. *International Journal of Psychology, 43*, 919-928.

Rebolj, M., van Ballegooijen, M., van Kemenade, F., Looman, C., Boer, R., & Habbema, J.D.F. (2008). No increased risk for cervical cancer after a broader definition of a negative Pap smear. *International Journal of Cancer, 123*, 2632–2635.

Rijmen, F., Vansteelandt, K., & De Boeck, P. (2008). Latent class models for diary method data: parameter estimation by local computations. *Psychometrika, 73*, 167-182.

Schepers, J., Ceulemans, E., & Van Mechelen, I. (2008). Selecting among multi-mode partitioning models of different complexities: A comparison of four model selection criteria. *Journal of Classification, 25*, 67-85.

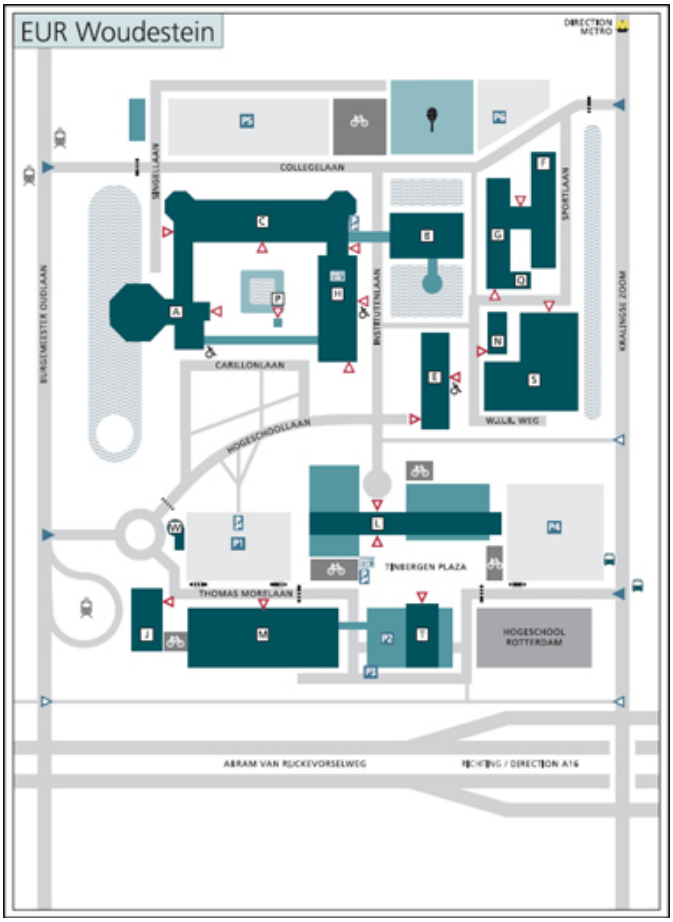
Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods, 40*, 61-72.

van der Velde, N., Meerding, W.J., Looman, C.W.N., Pols, H.A.P., & van der Cammen, T.J.M. (2008). Cost effectiveness of withdrawal of fall-risk-increasing drugs in geriatric outpatients. *Drugs & Aging, 25*, 521-529.

van Meerten, E., van der Gaast, A., Looman, C.W.N., Tilanus, H.W.G., Muller, K., & Essink-Bot, M.L. (2008). Quality of life during neoadjuvant treatment and after surgery for resectable esophageal carcinoma. *International Journal of Radiation Oncology, Biology, Physics, 71*, 160-166.

Route description

The Fall Meeting takes place at the Erasmus Universiteit Rotterdam, Campus Woudestein, T-Gebouw (T3-24), Burgemeester Oudlaan 50, 3062 PA Rotterdam.



Route description to campus Woudestein by Car

- The Woudestein campus is situated very near the Van Brienenoordbrug ('brug' = bridge). The Van Brienenoordbrug is crossed by the A16 motorway. This is where you have to get off the A16.
- If you come from the south (A16, direction Breda and Antwerp): keep following the signs Den Haag until you see the signs Capelle a/d IJssel and Rotterdam Centrum. Keep following the direction Rotterdam Centrum and you will see the direction to the university.
- If you come from the north (A13): keep following the direction Dordrecht/A16 until you see the sign Capelle a/d IJssel and Rotterdam Centrum. Keep following the direction Rotterdam Centrum and automatically you will see the direction of the university.
- If you come from the east (A12/A20): same route as coming from the north (Dordrecht/A16 and Rotterdam Centrum).

Route description to campus Woudestein by Public transport from NS Railway Station

- From Rotterdam Central Station

- with tram 21 in direction of De Esch. Get off at stop Woudestein.
- with tram 7 in direction Woudestein or Burg. Oudlaan. Get off at terminus.
- with metro in direction of Spijkenisse/Slinge. Change at station Beurs on metro in direction of Capelle a/d IJssel, Ommoord or Nesselande. Get off at stop Kralingse Zoom.

- From station Rotterdam Alexander

- with metro in direction of Schiedam Centrum/Spijkenisse. Get off at stop Kralingse Zoom.