

Sociology 952

EVENT HISTORY ANALYSIS

Syllabus – Fall 2019 (09/5/19 Version)

PREREQUISITES: Soc 361 and Soc 362 or the equivalents. Soc 674 or the equivalent is highly recommended. These prerequisites amount to a good working knowledge of the general linear model, models for binary and ordered dependent variables, and life tables. Familiarity with the properties of logarithms and exponentials is essential. Also, you should be comfortable using one of the major statistical software packages, preferably Stata as this is what we will use in class and is the package I will be best able to provide assistance with.

TIME AND PLACE: Monday 9:30-12:00pm, Social Science 6112

INSTRUCTOR: Christine Schwartz, 4458 Sewell Social Science Building, office: 262-5791, email: cschwartz@ssc.wisc.edu

Office Hours: Monday 1:30-2:30pm and by appt.

CLASS WEBSITE: I will use Canvas to post all materials for class including lectures, problems sets, data sets, and instructions for the final project. Reading will be posted on the website when it is not included in the assigned textbooks or available online through a Google Scholar “Find It at UW Madison” search.

SUBJECT MATTER: This course introduces regression analysis in which the outcome is time to event. Topics include review of life tables, discrete-time event history models, piecewise hazard models, proportional hazards, nonproportional hazards, parametric survival models, competing risks, heterogeneity, nested data, and other topics. If time is short, I will eliminate sections starting from the end. This is an “applied” course in the sense that the goal is to help you learn to carry out valid data analyses when the response is, or should be, time to event.

TEXTS: The following should be purchased:

Cleves, Mario A., William W. Gould, and Roberto G. Gutierrez. 2016. *An Introduction to Survival Analysis Using Stata, Revised Third Edition*. College Station, Texas: Stata Press. ISBN-10: 1-59718-174-9.*

Allison, Paul D. 2010. *Survival Analysis Using the SAS System: A Practical Guide, Second Edition*. Cary, NC: SAS Institute. ISBN 978-1-59994-640-5.**

*Note: The third edition of the Cleves et al. may also be purchased. There are two main differences between the revised third edition (2016) and the third edition (2010). (1) The revised edition has been updated to reflect Stata 14 code. (2) The revised edition contains a new section on predictive margins and marginal effects. These are both improvements and therefore the most recent edition is preferable, but I will cover marginal effects in class so it is not strictly necessary to buy the new edition.

**Note: You may also purchase the first edition of the Allison text. This is especially true if you are not using SAS as your main statistical package. Again, the most recent edition is preferable because it does contain some new material, but it is not strictly necessary to buy the new edition.

COMPUTATION: Almost all of the models discussed in this class can be estimated in Stata, with which students should already be familiar. Students who prefer to perform calculations using other software (R, SAS, etc.) are welcome to do so, but lecture examples will be done in Stata. Students who are unfamiliar with Stata may find it valuable to visit the Social Science Computing Cooperative's (SSCC) informative web site at <http://www.ssc.wisc.edu/sscc/pubs/stat.htm>. SSCC also holds periodic Stata training. Their calendar of classes can be found here: http://www.ssc.wisc.edu/sscc_jsp/training/index.jsp.

ASSIGNMENTS: The main requirement is a short empirical research paper that applies some of the methods covered in class to a substantive problem. The paper should include problem formulation, linkage to substantive literature, analysis of suitable data, and discussion of findings. Purely methodological papers will not be accepted. I recommend that students submit a two-page proposal that outlines the problem, data, and empirical analyses. Deadlines are:

Proposal: October 21

Final Draft: December 17

There will also be occasional homework assignments that should be completed by groups of roughly three in size. To obtain credit for homework, students must hand it in on time and be prepared to present solutions orally in class.

LEARNING OBJECTIVES: This course covers the following Sociology Graduate Program learning objectives:

1. Students will complete an original research project in one of the subfields of sociology.
2. Students will evaluate social science literature and employ most appropriate methods and practices in their own research.
3. Students will develop analytical thinking skills that enable them to evaluate information pertinent to their research question.
4. Students will communicate in a clear, organized engaging manner, using language, methods, and critical tools appropriate to the social sciences.

FORMAT: Lectures with short student presentations and discussion.

GRADING: The term paper determines 70 percent of the final grade. Homework and class participation determine 30 percent. **NO INCOMPLETES WILL BE GIVEN.**

SUPPLEMENTARY READING:

General Readings in Applied Social Statistics

Agresti, Alan . 2002. *Categorical Data Analysis* (Second Edition). New York. Wiley.

Cameron, A. Colin, and Pravin K Trivedi. 1998. *Regression Analysis of Count Data*.

Cambridge: Cambridge University Press.

Collett, D. 1991. *Modelling Binary Data*. London: Chapman and Hall.

Diggle, Peter J., Kung-Yee Liang, and Scott L. Zeger. 1994. *Analysis of Longitudinal Data*.

Oxford: Clarendon Press.

Greene, William H. 2002. *Econometric Analysis* (5th Ed.) Upper Saddle River, NJ: Prentice Hall.

Heinen, Tom. 1996. *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Thousand Oaks, CA: Sage Publications.

Hosmer, David W., and Stanley Lemeshow. 1989. *Applied Logistic Regression*. New York: Wiley.

Hsiao, Cheng. 1986. *Analysis of Panel Data*. Cambridge, UK.: Cambridge University Press.

- McCullagh, P. and J. A. Nelder. 1989. *Generalized Linear Models*. (2nd Ed.) London: Chapman and Hall.
- Powers, Daniel A. and Yu Xie. 2008. *Statistical Methods for Categorical Data Analysis* (Second Edition). Bingley, United Kingdom: Emerald Group Publishing.

Event History Analysis (with annotations by William M. Mason)

- Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*. Sage University Paper 46: Thousand Oaks, CA: Sage Publications. [A concise, beautifully written, nonmathematical introduction. Dated but still useful.]
- Blossfeld, Hans-Peter, Alfred Hamerle, and Karl Ulrich Mayer. 1989. *Event History Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates. [“Advanced” introduction; assumes comfort with calculus.]
- Blossfeld, Hans-Peter, and Götz Rohwer. 2002. *Techniques of Event History Modeling, Second Edition*. Mahway, NJ: Lawrence Erlbaum Associates. ISBN 0-8058-4091-5. [This book is also an introduction to the Rohwer-Pötter freeware for survival analysis named TDA.]
- Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004 *Event History Modeling: A Guide for Social Scientists*. Cambridge University Press. [This book was once required in this course along with the Cleves et al. and Allison. It is great.]
- Collett, David. 2003. *Modelling Survival Data in Medical Research, Second Edition*. New York: Chapman and Hall. [Collett is an exceptional expositor.]
- Collett, David. 2003. *Modelling Binary Data, Second Edition*. Boca Raton, FL: Chapman & Hall/CRC. [Superb book; oriented to the health sciences.]
- Cox, David R. and David Oakes. 1984. *Analysis of Survival Data*. New York: Chapman and Hall. [One of the statistical greats commits his textbook thoughts with a co-author. Compact; written for statistics graduate students and statisticians.]
- Hosmer, David W., Jr. and Stanley Lemeshow. 1999. *Applied Survival Analysis*. New York: Wiley. ISBN 0-471-15410-5. [Oriented to the health sciences; uses Stata for its examples; emphasis on proportional hazards model; worth investing in for its introductory expositions of martingales and counting processes.]
- Jenkins, Stephen P. 2005. *Survival Analysis*. [A highly useful manuscript written from an economist’s perspective. The mathematical level is higher than that of the assigned texts. See the course outline for the link. This text will be useful reading for those desiring more mathematical content than provided by the required texts.]
- Kalbfleisch, John D., and Ross L. Prentice. 2002 [1980]. *The Statistical Analysis of Failure Time Data, Second Edition*. New York: Wiley. [The classic treatise. Heavily revised and updated. Advanced.]
- Lancaster, T. 1990. *The Econometric Analysis of Transition Data*. New York: Cambridge University Press. [A keystone in the econometric literature on survival analysis. Advanced.]
- Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications. [Excellent introduction to the subject; detailed.]
- Long, J. Scott and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata, Second Edition*. College Station, TX: Stata Press. [Useful for its integration with Stata. However, this volume is not the equivalent of Allison's seemingly parallel text for SAS.]
- Preston, Samuel H., Patrick Heuveline, and Michel Guillot. 2001. *Demography*. Oxford: Blackwell Publishers. [Relevant here for its treatment of single- and multiple-decrement life tables.]
- Singer, Judith D. and John B. Willett. 2003. *Applied Longitudinal Data Analysis*. New York: Oxford University Press. [Singer and Willett are masterful expositors. The orientation in

their articles and in this text is “data-analytic, not theoretical.” The examples are all social science, and the assumed reader is an able student who has yet to invest heavily in math and statistics, yet is interested in the material. Part II (half the book, about 300 pages) is devoted to event history analysis.]

Therneau, Terry M. and Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag. [Therneau has written major survival analysis routines for R/S-Plus and SAS. The R/SPlus routines allow a Cox regression to be estimated using both stratification and a frailty term. This unusual flexibility allows the user to fit a hierarchically cross-classified survival model.]

Trussell, James, Richard Hankinson and Judith Tilton. 1992. *Demographic Applications of Event History Analysis*. New York: Oxford University Press. [Substantive articles as well as methodological—of interest to social scientists generally.]

Tuma, Nancy Brandon, and Michael T. Hannan. 1984. *Social Dynamics: Models and Methods*. Orlando, FL: Academic Press. [The first treatise on the subject by sociologists; rich and discursive.]

Vermunt, Jeroen K. 1997. *Log-Linear Models for Event Histories*. Thousand Oaks, CA: Sage Publications. [Chapters 4 and 5 integrate event history analysis with latent variables and missing data. Much useful exposition.]

SCHEDULE AND READINGS:

Except for the required texts, all course materials are available online. Unless otherwise indicated, the readings can be found via Google Scholar “Find it at UW Madison.” Readings located on the course website are shown as “on Learn@UW.” Both the weekly content and the pace of coverage shown here are approximate and subject to change. Readings should be completed prior to the class with which they are associated.

Week 1 (9/9): Introduction; administration; review; example.

Week 2 (9/16): Basic concepts of event history analysis.

Required:

Cleves, Gould, Gutierrez (2004)—hereafter CGG, ch. 1-6. (pp. 1-71. Chapters 1-4 are substantive. Chapters 5-6 are software-specific. The -stset- command (ch. 6) requires repeated study. Now is a good time to start.)

Allison (1995)—hereafter Allison, ch. 1-2

Optional:

Box-Steffensmeier & Jones (2004)—hereafter BSJ, ch. 1-2 (pp. 1-20)

Jenkins, Stephen P. 2005. *Survival Analysis*—hereafter “Jenkins,” ch. 1-2 (pp. 1-24).

[<http://www.iser.essex.ac.uk/files/teaching/stephenj/ec968/pdfs/ec968lnotesv6.pdf>]

Singer, Judith D., and John B. Willett. 1991. “Modeling the Days of Our Lives: Using Survival Analysis when Designing and Analyzing Studies of Duration and the Timing of Events.” *Psychological Bulletin*, Vol. 110(2), pp. 268-290.

Willett, John B., and Judith D. Singer. 1991. “From Whether to When: New Methods of Studying Student Dropout and Teacher Attrition.” *Review of Educational Research*, Vol. 61, pp. 407-450.

Week 3 (9/23): Estimating and comparing survival curves.*Required:*

CGG, ch. 7-8 (pp.73-119)

Allison, ch. 3

Optional:

Jenkins, ch. 4 (pp. 55-60)

Kaplan, Edward L. and Paul Meier. 1958. "Nonparametric Estimation from Incomplete Observations." *Journal of the American Statistical Association*, Vol. 53, pp. 457-481.

Week 4 (9/30): Lab session 1: Data management and descriptive analyses.*Required:*

Sweeney, Megan. 2002. Two Decades of Family Change: The Shifting Economic Foundations of Marriage." *American Sociological Review*

Week 5 (10/7): Discrete time, ties.*Required:*

Allison, ch. 7

Optional:

BSJ, ch. 5 (pp. 69-83)

Jenkins, ch. 6 (pp. 71-76)

Allison, Paul. 1982. "Discrete-Time Methods for the Analysis of Event Histories." In *Sociological Methodology 1982*, ed. S. Leinhardt, San Francisco, CA: Jossey-Bass, pp. 61-98.

Raftery, Adrian E., Steven M. Lewis, Akbar Aghajanian, and Michael J. Kahn. 1996.

"Event History Modeling of World Fertility Survey Data." *Mathematical Population Studies*, Vol. 6(2), pp. 129-153. (on Learn@UW)

Raftery, Adrian E., Steven M. Lewis, and Akbar Aghajanian. 1995. "Demand or Ideation? Evidence from the Iranian Marital Fertility Decline." *Demography*, Vol. 32(2), pp. 159-182.

Lewis, Steven M. and Adrian E. Raftery. 1999. "Bayesian Analysis of Event History Models with Unobserved Heterogeneity via Markov Chain Monte Carlo." *Sociological Methods & Research*, Vol. 28, pp. 35-60.

Note: The three Raftery *et al.* articles also discuss heterogeneity.

Singer, Judith D., and John B. Willett. 1993. "It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events." *Journal of Educational Statistics*, Vol. 18(2), pp. 155-195.

Jenkins, Stephen P. 1997. "Discrete Time Proportional Hazards Regression." *Stata Technical Bulletin*, STB-39, pp. 22-32.

Jenkins, Stephen P. 1995. "Easy Estimation Methods for Discrete-Time Duration Models." *Oxford Bulletin of Economics and Statistics*, Vol. 57, pp. 129-138.

Holford, Theodore R. 1976. "Life Tables with Concomitant Information." *Biometrics*, Vol. 32, pp. 587-597.

Prentice, R.L. and L.A. Gloeckler. 1978. "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data." *Biometrics*, Vol. 34, pp. 57-67.

Doksum, Kjell A. and Miriam Gasko. 1990. "On a Correspondence between Models in Binary Regression Analysis and in Survival Analysis." *International Statistical Review*, Vol. 58, pp. 243-252. (M)

Firth, David and Clive Payne, Joan Payne. 1999. "Efficacy of Programmes for the Unemployed: Discrete Time Modelling of Duration Data from a Matched

Comparison Study." *Journal of the Royal Statistical Society, Series A*. Vol. 162, pp. 111-120.

for background reading on categorical data analysis, esp. on the logit model:
 Long, J. Scott (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage, chapters 3-4. (on Learn@UW)

Week 6 (10/14): Lab session 2: Regression modeling – discrete-time methods

Week 7 (10/21): Cox regression. (PAPER PROPOSAL DUE)

Required:

CGG, ch. 9-11 (pp. 175-195)

Allison, ch. 5

Optional:

BSJ, ch. 4 (pp. 47-67), ch. 7 (pp. 95-117), ch 8. (pp. 119-137)

Jenkins, ch. 5 (pp. 61-70), ch. 7 (pp. 77-80)

Piliavin, Irving, and Bradley R. Entner Wright, Robert D. Mare, and Alex H. Westerfelt. 1996. "Exits from and Returns to Homelessness." *Social Service Review*, Vol. 70(1), pp. 33-57.

Royston, Patrick (2001). "Flexible Parametric Alternatives to the Cox Model, and More." *The Stata Journal*, Vol. 1, pp. 1-28.

Royston, Patrick and M.K.B. Parmar. (2002). "Flexible Parametric Models for Censored Survival Data, with Application to Prognostic Modelling and Estimation of Treatment Effects." *Statistics in Medicine*, Vol. 21, pp. 2175-2197.

Cox, David R. (1972). "Regression Models and Life Tables," *Journal of the Royal Statistical Society, Series B*, Vol. 34, pp. 187-220 (with discussion).

Mencken, Jane, James Trussell, Debra Stempel, and Ozer Babakol. 1981. "Proportional Hazards Life Table Models: An Illustrative Analysis of Socio-Demographic Influences on Marriage Dissolution in the United States." *Demography*, Vol. 18(2), pp. 181-240.

Week 8 (10/28): Lab session 3: Regression modeling – Cox regression

Week 9 (11/4): Parametric and piecewise parametric models.

Required:

CGG, ch. 12-14

Allison, ch. 4

Optional:

BSJ, ch. 3 (pp. 21-46), ch. 8 (pp. 137-139)

Jenkins, ch. 3 (pp. 25-54)

Bruderl, J. and A. Diekmann. 1995. "The log-logistic Rate Model." *Sociological Methods & Research*, Vol. 24, pp. 158-186.

Defo, Barthelemy Kuate. 1997. "Effects of Infant Feeding Practices and Birth Spacing on Infant and Child Survival: A Reassessment from Retrospective and Prospective Data." *Journal of Biosocial Science*, Vol. 29, pp. 303-326.

Defo, Barthelemy Kuate and Alberto Palloni. 1995. "Determinants of Mortality Among Cameroonian Children: Are the Effects of Breastfeeding and Pace of Childbearing Artifacts?" *Genus*, Vol. 51(3-4), pp. 61-96.

Diekmann, Andreas and Henriette Engelhardt. 1999. "The Social Inheritance of Divorce: Effects of Parent's Family Type in Postwar Germany." *American Sociological Review*, Vol. 64. pp. 783-793.

Week 11 (11/11): Lab session 4: Regression modeling – Parametric methods

Week 9 (11/18): Competing risks.

Required:

Allison, ch. 6,

Optional:

Jenkins, ch. 9 (pp. 91-112)

Tutz, Gerhard. 1995. "Competing Risks Models in Discrete Time with Nominal or Ordinal Categories of Response." *Quality & Quantity*, Vol. 29, pp. 405-420.

Hachen, David S., Jr. 1988. "The Competing Risks Model." *Sociological Methods & Research*, Vol. 17, pp. 21-54.

Week 12 (11/25): Heterogeneity, repeated events.

Required:

CGG, ch. 15 (pp. 269-298)

Allison, ch.8

Optional:

BSJ, ch. 9-10 (pp. 141-182)

Jenkins, ch. 8 (pp. 81-90)

Trussell, James and German Rodriguez. 1990. "Heterogeneity in Demographic Research." Pp. 111-132 in Julian Adams, David A. Lam, Albert I. Hermalin and Peter E. Smouse (eds.), *Convergent Issues in Genetics and Demography*. New York: Oxford University Press. (on Learn@UW)

Hoem, Jan M. 1990. "Identifiability in Hazard Models with Unobserved Heterogeneity: The Compatibility of two Apparent Contradictory Results." *Theoretical Population Biology*, Vol. 37, pp. 124-128.

Pickles, Andrew and Robert Crouchley. 1995. "A Comparison of Frailty Models for Multivariate Survival Data." *Statistics in Medicine*, Vol. 14, pp. 1447-1461.

Gutierrez, Roberto G. 2002. "Parametric Frailty and Shared Frailty Survival Models." *The Stata Journal*, Vol. 2, pp. 22-44.

Blossfeld, Hans-Peter, and Alfred Hamerle. 1992. "Unobserved Heterogeneity in Event History Models." *Quality & Quantity*, Vol. 26, pp. 157-168.

Allison, Paul. 1996. "Fixed-Effects Partial Likelihood for Repeated Events." *Sociological Methods & Research*, Vol. 25(2), pp. 207-22.

Willett, John B., and Judith D. Singer. 1995. "It's Deja Vu All Over Again: Using Multiple-Spell Discrete-Time Survival Analysis." *Journal of Educational and Behavioral Statistics* Vol. 20, pp. 41-67.

Steele, Fiona, Ian Diamond, and Sajeda Amin. 1996. "Immunization Uptake in Rural Bangladesh: a Multilevel Analysis." *Journal of the Royal Statistical Society, Series A*, Vol. 159, part 2, pp. 289-299.

Guo, Guang, and German Rodriguez. 1992. "Estimating a Multivariate Proportional Hazards Model for Clustered Data Using the EM Algorithm, with an Application to child Survival in Guatemala." *Journal of the American Statistical Association*, Vol. 87, pp. 969-76.

Guo, Guang. 1993. "Use of Sibling Data to Estimate Family Mortality Effects in Guatemala." *Demography*, Vol. 30(1), pp. 15-32.

- Barber, Jennifer S., Susan A. Murphy, William G. Axinn, and Jerry Maples (2001).
Discrete-Time Multilevel Hazard Analysis. *Sociological Methodology* 31: 201-235.
- Heckman, James J., and James R. Walker. 1990. "The Third Birth in Sweden." *Journal of Population Economics* 3: 235-75.

Week 13 (12/2): Lab session 5: Competing risks, heterogeneity, repeated events

Week 14 (12/9): Miscellaneous, Reports on Student Work in Progress.

Required:

Allison, ch. 9

Optional:

BSJ, ch. 6 (pp. 85-93), ch. 11 (pp. 183-198)

Hoem, Jan M. 2000. "Systematic Patterns of Zero Exposures in Event-History Analysis."
Sociological Methodology, Vol. 30, pp. 237-259.

Yamaguchi, Kazuo. 1990. "Logit and Multinomial Models for Discrete-Time Event-History Analysis: A Causal Analysis of Interdependent Discrete-State Processes."
Quality and Quantity, Vol. 24., pp. 323-341.

Petersen, Trond. 1995. "Models for Interdependent Event-History Data: Specification and Estimation." *Sociological Methodology*, Vol. 25, pp. 317-375.

FINAL PAPERS ARE DUE TO CANVAS BY 5:00 P.M. ON TUESDAY, DECEMBER 17.