

SW 832: Research Methods for Social Policy Analysis  
Fall 2014  
Focus on Multilevel and Longitudinal Data Analysis

"[We] should be aware of state of the art treatment for specific disorders as represented in the scientific literature and endorsed by government health agencies. If the worker cannot treat with specificity, then the client should be referred for proper treatment. The impression should not be given that all treatments are equal or that treatments of demonstrated efficacy are simply alternatives. (p11)"

— Mary Richmond as recounted in Myers and Thyer (1997) "Should Social Work Clients Have the Right to Effective Treatment?"

"Suppose you are suffering from a distressing illness, painful or incapacitating, and your physician says that it would be a good idea to have surgeon X perform a certain radical operation in the hope of curing you. You would naturally inquire whether this operation works for this disease and how risky it is. The physician might say, "Well, it doesn't always work, but it's a pretty good operation. It does have some risk. There are people who die on the operating table, but not usually." You would ask, "Well, what percentage of times does it work? Does it work over half the time, or 90%, or what? And how many people die under the knife? One in a thousand? If it were five in a hundred, I don't know that I'd want to take the chance, even though this illness is irksome to me." How would you react if your physician replied, "Why are you asking me about statistics? We are talking about you—an individual patient. You are unique. Nobody is exactly like you. Do you want to be a mere statistic? What differences do those percentages make, anyway?" We do not think a person should be pleased if the doctor replied in that evasive fashion. Why not? Because, as Bishop Butler (1736) said, probability is the guide of life."

— Grove and Meehl (1996) in "Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy"

"We may at once admit that any inference from the particular to the general must be attended with some degree of uncertainty, but this is not the same as to admit that such inference cannot be absolutely rigorous, for the nature and degree of the uncertainty may itself be capable of rigorous expression."

— Sir Ronald A. Fisher

"Mathematics is the art of giving the same name to different things."

— Henri Poincare

"The best thing about being a statistician is that you get to play in everyone's backyard."

— John Tukey

**Credit Hours:** 3  
**Prerequisites:** None  
**Instructor:** Andrew Grogan-Kaylor, MA, MSSW, PhD  
**Office:** 3847 School of Social Work

**Phone:** 615-3369  
**Email:** [agrogan@umich.edu](mailto:agrogan@umich.edu) (best way to get in touch with me)  
(please put 832 in the subject line)  
**Office Hours:** Wednesday 9-11 or by appointment

Course Web site at <http://ctools.umich.edu/>

## **Course Description and Content**

Multilevel models have become a standard statistical tool for quantitative research on neighborhoods, communities and schools. Perhaps surprisingly, the multilevel model for cross-sectional data can easily accommodate longitudinal data where study participants are observed repeatedly. While this is sometimes not recognized, multilevel models for longitudinal data are closely related to other important longitudinal data models, such as fixed effects regression, an important technique for controlling for unobserved variables.

This course focuses on the use of multilevel and longitudinal data analysis for social work research. The course is conceptualized as covering the following topics:

- 1) The multilevel model for cross-sectional data.
- 2) The extension of multilevel modeling to repeated measures in longitudinal research (i.e. growth trajectory models).
- 3) Possible additional topics based on student interest and availability of time.
  - a. Models for dichotomous outcomes.
  - b. Other panel data models such as fixed effects and random effects models.
  - c. The multilevel model for meta-analysis
  - d. Models for "cross-classified" nestings.

Although there are multilevel models for dichotomous outcomes, the models discussed for the majority of this course are primarily intended, and best developed, for continuous outcomes, so are most appropriate for students with research questions that can in some way be conceptualized as continuous outcomes. The course is intended for students who have some familiarity with ordinary least squares regression, since the OLS model will be used as a starting point for discussion.

The treatment of these topics involves some discussion of the underlying statistical theory, but is much more focused on the application of these models with real data. STATA is intended to be the primary statistical software used in this course. Other software packages may be discussed as student interest suggests, and time allows.

## **Method of Instruction**

Instruction is conceived as a mixture of in class lecture and discussion and time spent in the computer lab working through actual data problems with statistical software in order to convey relevant content.

## **Required Readings**

My presentation of the material for the course will draw on articles and chapters from several different books. Given the high price of many statistical books, there is no one book that I am drawing on consistently enough to justify the purchase of a text.

Therefore, all readings for this course will be made available using CTools.

## Course Requirements

- 1) **Two 8-10 page “mini-papers” focusing upon the statistical analysis on a research question of interest using secondary or other available data.** The first of these papers is intended to be a cross-sectional multilevel analysis and the second paper is to be a longitudinal repeated measures analysis. At a minimum, each paper should include some basic demographic controls in addition to one or two independent variables of theoretical or substantive interest. At times, your demographic variables may in fact be the relevant independent variables of theoretical interest. The two papers will be due at different points over the term. Consultation with the professor and other students will be available. Quality of writing, and clarity of presentation, will be one of the criteria used in grading. Please make sure that you cite other’s work properly and avoid plagiarism<sup>1</sup>—when discovered—will be dealt with severely. Each paper should include: a review of at least 5 key pieces of background literature<sup>2</sup>; descriptive statistics; a multivariate multilevel analysis<sup>3</sup>; a well-documented and nicely formatted and labeled graph illustrating some bivariate aspect of your data; and a discussion of your findings and their implications. NB: The analysis in each paper should include testing and discussion of the meaning of a random intercept, and at least 2 random slope components, even though your final model does not need to have either a random intercept or random slope. Further details of these assignments will be forthcoming and will also appear on the class Web site.
  - a. **Cross-Sectional Multilevel Mini-Paper.** How are multiple factors associated with a continuous outcome of interest? This paper should make use of data appropriate for multilevel analysis, generally data containing information on individual subjects nested inside neighborhoods regions or schools that contains some type of identifier (often anonymous) of the neighborhood, region or school. Please note that because of the way our presentation in this course is structured, it is easiest to focus on a continuous outcome. **[30% of grade]**
  - b. **Longitudinal Repeated Measures Mini-Paper.** How are multiple factors associated with a continuous outcome of interest? This paper should make use of data appropriate for a longitudinal, repeated measures, analysis, based in the multilevel tradition. Data that are appropriate for this kind of analysis will contain repeated observations on multiple subjects and are often called *panel* data, that contains some type of identifier (often anonymous) of each individual in the study. Please note that because of the way our presentation in this course is structured, it is easiest to focus on a continuous outcome. **[30% of grade]**
- 2) **Class participation.** You will learn more the more you participate. Because this is a doctoral course it is important to participate actively, and especially to ask questions. I recognize that there

---

<sup>1</sup> Please note that for purposes of this course, plagiarism consists of six or more consecutive words, taken from another source without proper attribution. Failure upon my part to detect plagiarism does not imply approval of plagiarism.

<sup>2</sup> Please note that this literature review is not intended to be exhaustive, merely to pave the way for the essential argument of the paper.

<sup>3</sup> While I have no set format for presentation of tables of results, please pay some attention to formatting your results, and do not simply paste unformatted statistical output directly into your paper.

are many variations that affect the rate of participation in this course, but a portion of your grade will depend on class participation. **[20% of grade]**

- 3) **Attendance.** You are expected to attend each class and lab session. If you miss a class session for any reason, please see me to make sure that you are grasping the material. In particular, you will be expected to complete all lab assignments, unless we make other arrangements.
- 4) **Lab Work.** We will spend some time every week doing lab work. Some computers are available via the SSW mobile computer lab<sup>4</sup>. The purpose of the lab is to give you hands on experience with the building blocks of multilevel and repeated measures longitudinal analysis as well as working with the specific kinds of data structures pertaining to this type of analysis. Most of the learning in lab occurs in the doing of lab. Lab is intended to be a collaborative endeavor where students work with the instructor and with each other. I will not directly grade your lab work, or be able to provide written feedback on your lab work. I am always happy to discuss lab work during lab, during office hours, or over e-mail.

A note on work handed in late: Most students turn in work in accordance with class deadlines. In order to be fair to the majority of students, I have developed the following policy: late work will be graded down by half a grade a day unless prior arrangements for an extension have been made with me. I very much understand that extenuating circumstances may arise which make it difficult to turn in work on time. All I am asking you to do is to communicate with me if you need some kind of extension so that we can work out an arrangement that is mutually agreeable.

### Grades

Class Participation	20%
Choosing First Data Set	10%
First Mini-Paper	30%
Choosing and Preparing Second Data Set	10%
Second Mini-Paper	30%

### Grading Scale<sup>5</sup>

"A" grades are given for exceptional individual performance and mastery of the material. The use of "A+", "A", and "A-" distinguishes the degree of superiority. "B" grades are given to students who demonstrate mastery of the material, at the level expected for the course. "B+" is used for students who perform just above the mastery level but not in an exceptional manner. "B-" is used for students just below the mastery level. "C" grades are given when mastery of the material is minimal. A "C-" is the lowest grade which carries credit. "D" grades indicate deficiency and carry no credit. "E" grades indicate failure and carry no credit.

### Grading Criteria for Written Assignments<sup>6</sup>

The two mini-papers will be graded using the following criteria:

---

<sup>4</sup> Instead of using the mobile lab, you may use your own laptop to connect to <http://virtuallsites.umich.edu>. NB that virtuallsites is much easier to use if your virtual desktop is the same size as your local desktop. Let me know if you need help with this.

<sup>5</sup> Adapted from the MSW student guide

<sup>6</sup> Adapted from a statement by Michael Reisch.

- 1) The quality of the analysis and depth of understanding of the concepts, ideas, and information presented.
- 2) The clarity of expression and organization of the paper – i.e., a logical order to the presentation of its major themes.
- 3) The appropriate use of references and resources, and the variety/range of resources referenced.
- 4) The use of proper grammar and the overall professional presentation of the paper.

**Note:** Please take time to proofread your writing and make sure you submit a second or third draft.

## **Data Sets**

Both of the written paper projects for this course require you to make use of previously collected data, usually publicly available secondary data, for your analyses. I will work with you to identify suitable data. Below are some suggested data sets for the two major types of analyses that we will discuss in the course.

It may be the case that you have access to data that is not publicly available, for example data from a research project, or agency, with which you are involved. It is perfectly acceptable to use such data, but you must have a copy of this data in hand at the time that you turn in your one paragraph on each study.

In general, and where concerns of access permit, I intend to post any publicly accessible data that you are using for your projects on CTools.

***Multilevel Cross-Sectional Analyses*** (data appropriate for these projects are generally data on individuals nested inside schools, neighborhoods or regions; such data have an identifier for the school, neighborhood or region in which the individuals are nested).

- The High School and Beyond Survey is a nationally representative survey of high school students suitable for multilevel analysis. While this is an older survey, it has the advantage of being the subject of statistical examples in many multilevel textbooks. If there is interest in this survey, I will post it on CTools.
- The National Youth Survey is a nationally representative survey of youth. Youth are nested in “census regions” which can be used as a nesting variable for this data. I have students work successfully with this data in the past, and it is posted on CTools.
- The Program on Human Development in Chicago Neighborhoods (PHDCN) is publicly available via <http://www.icpsr.org>. I have not worked with this data set before, but understand from colleagues that it is appropriate for multilevel analysis. There is a public use version available, and if there is student interest, we can obtain this data for class. I understand that extracting PHDCN variables is somewhat difficult, so working with this data may involve extra effort on your part.
- “The National Longitudinal Study of Adolescent Health (Add Health) is a nationally representative study that explores the causes of health-related behaviors of adolescents in grades 7 through 12 and their outcomes in young adulthood. Add Health seeks to examine how social contexts (families, friends, peers, schools, neighborhoods, and communities) influence adolescents’ health and risk behaviors.” (from <http://www.cpc.unc.edu/addhealth>). I have not worked with this data set before, but understand from colleagues that it is appropriate for multilevel analysis. There is a public use version available, and if there is student interest, and

the cost is reasonable, we can obtain this data for class. (The server with pricing information has been down recently; perhaps this is not a good sign).

**Longitudinal Repeated Measures Analyses** (*data appropriate for these projects are generally data on individuals observed repeatedly over time; such data have an identifier for each individual*).

- A very useful data set for these kinds of analyses is the National Longitudinal Survey of Youth, a database that I have used in my previous research. Since I am familiar with this database I am more likely to be able to help you answer questions about its use. I have posted a brief introduction to the NLSY, created for other courses, at [http://www-personal.umich.edu/~agrogan/NLSY/nlsy\\_basic\\_info.htm](http://www-personal.umich.edu/~agrogan/NLSY/nlsy_basic_info.htm). In previous courses I have taught using the NLSY, the applications have been mostly cross-sectional and indeed a mostly cross-sectional extract of the NLSY is posted in <http://www-personal.umich.edu/~agrogan/NLSY/>. However, for this course, you will need to create a longitudinal data set on your own. I will work with you to get your data appropriately formatted for your project. However, as this is a doctoral course, I think that it is a valuable learning exercise for you to do much of this work yourself, although I am certainly happy to offer consultation and advice.

**Other Possible Sources Of Data** (*If you locate a **suitably structured** data set via one of these other sources, I will work with you to get it appropriately formatted for your project. However, as this is a doctoral course, I think that it is a valuable learning exercise for you to do much of this work yourself, although I am certainly happy to offer consultation and advice.*)

- ICPSR (<http://www.icpsr.org>), the campus data warehouse, is a worthwhile place to search for suitably structured data sets that fit your own research interests.
- The Social Science Data Analysis Network (<http://www.ssdan.net/>) is a resource on campus of which I have only recently become aware, but may also be a source for Census related data sets.
- The National Data Archive On Child Abuse And Neglect at Cornell University (<http://www.ndacan.cornell.edu/>) is a possible source for child welfare related data sets. I have worked with several NDACAN data sets in my own research, particularly the NSCAW.

### **Software (you may need to purchase software for this course)**

- STATA (available in most campus computer labs and via virtualsites) (in class exercises will focus on the use of STATA although the statistical concepts covered transcend any one statistical software package)
- We will make some small amount of use of the HLM software developed by Raudenbush and Bryk. SITES computer labs do not currently host HLM, therefore, if we use this software, we will need to download the student version to student laptops from <http://www.ssicentral.com/hlm/student.html>.
- If time and interest permit, we may make some use of SAS Proc Mixed as well as R procedures for multilevel modeling to conduct multilevel analysis.

### **Suggested Ancillary Texts**

Although all course readings are available on CTools, you may find some of the following texts to be useful in your study of multilevel models. **None of the following texts are required**; they are simply

listed as an aid in your education. Should you have questions, I am happy to discuss what I see as the merits of various texts at any time.

The classic text on HLM is:

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

In my mind the Raudenbush and Bryk text remains one of the most pedagogically accessible introductory texts on hierarchical or multilevel models.

The HLM Program Manual is also a very accessible guide to HLM.

Raudenbush, S. W., Bryk, A. S. & Congdon, R. (2001). *HLM: Hierarchical linear and nonlinear modeling*.

For longitudinal applications of multilevel modeling, the best introductory text is the first half of:

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford ; New York : Oxford University Press.

Lastly, as one moves forward in one's understanding of multilevel modeling, certain technical questions sometimes arise. I've found the best sources of answers to very technical questions to be the Raudenbush and Bryk text, or the recent book by Rabe-Hesketh and Skrondal. This latter book seems to be less helpful for beginners, however.

Rabe-Hesketh, S. & Skrondal (2012) *Multilevel and Longitudinal Modeling Using Stata (Third Edition)*. College Station, TX: Stata Press.

Week	Date	Topic	Readings	Lab	Assignments
1	September 3 <sup>rd</sup> , 2014	welcome and introduction to the class	<p>Please read the following article as a general overview of the way in which neighborhood research interacts with the use of multilevel modeling.</p> <p>Holland, S., Burgess, S., Grogan-Kaylor, A., &amp; Delva, J. (2010). Understanding neighbourhoods, communities and environments: New approaches for social work research. <u>British Journal of Social Work</u>.</p>		
2	September 10 <sup>th</sup> , 2014	<p>Review of central limit theorem, classical statistical inference &amp; t test, OLS including:</p> <ul style="list-style-type: none"> <li>• How graph of regression relationship relates to regression parameters</li> <li>• interpretation of <math>\beta</math>s, se's, t's and p's</li> <li>• <math>R^2</math></li> <li>• dummy variables</li> <li>• interactions</li> <li>• OLS regression assumptions</li> </ul>	<p>Chs. 1 &amp; 4 from Hamilton, L. C. (2006). <u>Statistics with STATA: Updated for version 9</u>. Belmont, CA: Thomson-Brooks/Cole.</p> <p>Grogan-Kaylor regression tutorial available on CTools.</p> <p>Stata regress documentation</p> <p>"Two Page Stata" available via CTools</p>	<p>introduction to Stata</p> <p>regression and regression graphics in Stata</p>	1 paragraph on cross-sectional multilevel analysis project; template to be provided on CTools



Week	Date	Topic	Readings	Lab	Assignments
3	September 17 <sup>th</sup> , 2014	introduction to multilevel (hierarchical) modeling: the multiple equation model	Chs. 2 & 4 from Raudenbush, S. W., & Bryk, A. S. (2002). <i>Hierarchical linear models: Applications and data analysis methods</i> . Thousand Oaks, CA: Sage Publications.	running simple multiple equation models in HLM	those w/ laptops install a student version of HLM available from <a href="http://www.ssicentral.com/hlm/student.html">http://www.ssicentral.com/hlm/student.html</a>
4	September 24 <sup>th</sup> , 2014	introduction to mlm single equation approach	Stata <i>mixed</i> documentation from StataCorp (2013). <i>Stata 13 Base Reference Manual</i> . College Station, TX: Stata Press.  Ciarleglio, & Makuch (2007). Hierarchical linear modeling: An overview. <i>Child abuse &amp; neglect</i> , 31(2), 91.	running simple single equation models in Stata	
5	October 1 <sup>st</sup> , 2014	introduction to mlm single equation approach (continued)	Merlo et al. (2005). A Brief Conceptual Tutorial on Multilevel Analysis in Social Epidemiology: Interpreting Neighbourhood Differences and the Effect of Neighbourhood Characteristics on Individual Health, <i>Journal of Epidemiology and Community Health</i> , 59(12), 1022-1028.		

Week	Date	Topic	Readings	Lab	Assignments
6	October 8 <sup>th</sup> , 2014	<b>Lab Day 1</b>	Aarts et al. (2014). A solution to dependency: using multilevel analysis to accommodate nested data, <i>Nature Neuroscience</i> , 17(4), 491-496.		
7	October 15 <sup>th</sup> , 2014	mlm for longitudinal applications	Review <i>mixed</i> documentation  Ch. 3 from Singer, J. D., & Willett, J. B. (2003). <i>Applied longitudinal data analysis : Modeling change and event occurrence</i> . Oxford ; New York : Oxford University Press.  p. 62-72 from Luke, D. A. (2004). <i>Multilevel modeling</i> . Thousand Oaks, CA : Sage.		first minipaper due

Week	Date	Topic	Readings	Lab	Assignments
8	October 22 <sup>nd</sup> , 2014	mlm for longitudinal applications 2	Delva, J., Grogan-Kaylor, A., Andrade, F., Hynes, M., Sanchez, N., & Bares, C. (Forthcoming). An agenda for Longitudinal Research on Substance Use and Abuse with Hispanics in the U.S. and with Latin American Populations. In Thomas, Y. F., Price, L.N. & Lybrand, A. V. (Eds.), <u>Drug Use Trajectories among African American and Hispanic Youth</u> . Springer Publishers.  (focus on section on longitudinal modeling)		1 paragraph on longitudinal repeated measures multilevel analysis project
9	October 29 <sup>th</sup> , 2014	mlm using SAS proc mixed or R packages  -or-  Advanced topics with error structures  -or-  Emails from colleagues with MLM application questions	Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. <i>Journal of Educational and Behavioral Statistics</i> , 24(4), 323-355.		
10	November 5 <sup>th</sup> , 2014	non-normal outcomes	meqrlogit documentation		

Week	Date	Topic	Readings	Lab	Assignments
11	November 12 <sup>th</sup> , 2014	Fixed effects models 1	Allison, P. (2006) Fixed Effects Regression Methods, <i>SAS Users Group International Proceedings</i> , Cary NC: SAS Institute.		
12	November 19 <sup>th</sup> , 2012	Fixed effects 2 or meta-analysis 1	TBD		
★★★ 13	November 26 <sup>th</sup> , 2014	Meta-analysis 1 or consultation day	TBD		
13	December 3 <sup>rd</sup> , 2012	Meta-analysis 2?	TBD		
14	December 11 <sup>th</sup> , 2012	<b>Lab Day 2</b>			second minipaper due December 18 <sup>th</sup> , 2012