

Linking longitudinal study site data and instruction in analysis: a replicable approach to improving public health training and research

Revision to manuscript BLT/2007/047357

Jill R. Williams,^{a,b} Enid J. Schatz,^{a,b,c} Benjamin D. Clark,^b Mark A. Collinson,^b Samuel J. Clark,^{a,b,d} Jane Menken,^{a,b} Kathleen Kahn^b, Stephen M. Tollman^b

Corresponding author:

Jill R. Williams

UCB 483

Boulder, Colorado 80309-0483

303-492-5253

Jill.Williams@Colorado.edu

^a Institute of Behavioral Science, Population Program, University of Colorado-Boulder, UCB 483, Boulder, CO, 80309.

^b MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences 10th Floor, University of the Witwatersrand, 7 York Road, Parktown 2193 South Africa,

^c University of Missouri-Columbia, School of Health Professions, 420 Lewis Hall, Columbia MO 65203

^d University of Washington, Department of Sociology 202 Savery Hall, Box 353340, Seattle WA 98195-3340

Linking longitudinal study site data and instruction in analysis: a replicable approach to improving public health training and research

Abstract

Problem Linking public health training programs to health research has been proposed as a mechanism for improving the quality of training programs and increasing the number of public health personnel in Africa. Opportunities exist in Africa and elsewhere to link training programs to existing demographic and health surveillance systems. However, difficulty in accessing data and lack of technical training in use of complex longitudinal data are obstacles to creating a sustainable link between research and training programs.

Approach Couple user-friendly data access with intensive technical training on longitudinal data management and analysis.

Local Setting The University of the Witwatersrand, South Africa, has established new masters and PhD programs in public health and population studies. The MRC/Wits Rural Public Health and Health Transitions Research Unit (formerly the Agincourt Health and Population Unit) in northeast South Africa is part of the Wits School of Public Health.

Relevant Changes Access to data was improved by creating a system for extracting anonymized fully-edited subsamples from the Agincourt Health and Demographic Surveillance System (AHDSS). Technical training was offered to students through an intensive short course on longitudinal data analysis that uses an AHDSS subsample. The course has been offered three times, 2006-2008, as a collaboration between Wits and the University of Colorado-Boulder.

Lessons Learned Early results demonstrate accelerated student training, improved student research, and increased use of the AHDSS by students and more experienced researchers. The approach is adaptable to other partnerships between schools of public health and health and demographic research sites.

Introduction

The *Bulletin of the WHO* recently drew attention to the shortage of health professionals in low- to middle-income countries.¹ The situation is especially severe in Africa, where public health education capacity is particularly low.² Few are trained to design, implement, and evaluate public health policies and programs and contribute to improvement of health information systems. Linking public health education to health research has been cited as a way to increase both the number and quality of public health professionals.²⁻⁵ Indeed, a growing number of schools of public health are connected to field sites in developing countries that carry out health and demographic surveillance, collecting data at least yearly.⁶ Trainee access to these longitudinal data can provide exceptional opportunities for students to learn first-hand about health in the populations they intend to serve.

However, connecting training programs to demographic surveillance sites poses challenges to both partners. Field sites may lack resources to support training, including data preparation, and must ensure that trainee datasets protect confidentiality of human subjects. Training programs need to provide students and supervisors the resources and technical skills to analyze complex longitudinal data. This paper describes our experience overcoming these challenges while

connecting public health graduate training at the University of the Witwatersrand (Wits), South Africa, to the Agincourt Health and Demographic Surveillance System (AHDSS). This effort is part of research and training collaborations between the Wits School of Public Health⁷ and its MRC/Wits Rural Public Health and Health Transitions Research Unit,⁸ and the University of Colorado-Boulder (CU) Institute of Behavioral Science,⁹ United States, and its African Population Studies Research and Training Program.¹⁰

Wits School of Public Health and the Agincourt Health and Demographic Surveillance System

The AHDSS was established by Wits School of Public Health faculty in rural north-east South Africa to provide population-based health and demographic information to inform the restructuring of the health care system in post-apartheid South Africa. It began with a baseline census in 1992 and continues with annual updates of household membership and vital events (births, deaths, migrations). It currently covers a sub-district of 25 villages, with some 14,800 households and a population of over 84,000 people. It provides reliable information on household and individual social and economic status and on causes of death (from verbal autopsies conducted on all reported deaths).¹¹

The AHDSS provides information of the breadth and depth necessary for understanding health and health behavior, monitoring health and population change, studying dynamics of health and socio-economic transition, preparing and conducting community trials, and evaluating health and development programs in its region.¹² Its data provide invaluable opportunities for training new Masters level (MSc, MPH, MA) and PhD researchers in the population and health science programs at Wits.

Initial efforts to involve students highlighted obstacles to gaining access to and then analyzing information from this complex dataset. Important barriers are discussed below, followed by a description of our efforts to overcome them.

Access to AHDSS Data

Until recently, the steps for obtaining AHDSS data were:

- a student or faculty supervisor submitted a proposal and data request;
- the Agincourt team reviewed the proposal for scientific merit, feasibility, and fit to the research program;
- after approval, an Agincourt staff member wrote unique extraction script to create a tailored dataset.

A significant challenge was that students and supervisors lacked familiarity with longitudinal data generally and AHDSS data specifically. This lack led to underspecified requests which the Agincourt staff found difficult to decipher. Consequently, time to approval and data preparation was often long, making it difficult for students to complete research within the time specified by their programs. A further challenge was the burden on staff of dataset preparation, leading to underutilization of the AHDSS by students and even experienced researchers.

In 2004, the authors collectively began reformulating the process of obtaining AHDSS data to include both technical training and the possibility of exploratory analysis with the data prior to requesting a tailored dataset. A crucial first step was creation of an extraction program that can

produce a sample database – from 10 to 100% sample – of some or all of the information in the AHDSS and with anonymized individual identifiers to protect human subject confidentiality. The sample database maintains the structure of the full relational Agincourt database. Its means fall within one standard deviation of AHDSS means for key variables; thus, results are expected to be comparable between sample and full Agincourt databases. User-friendly documentation, including descriptions of the study setting and database structure, a data dictionary, and a standard user-agreement, were prepared.

Using the extraction program, the Agincourt Anonymized 10% Sample Database (AS-10) was created. The AS-10 includes 10% of the geographic locations in each village where individuals and social groups reside and information on all individuals in each location over the full period of data collection. The AS-10 is updated annually.

Wits has made the 2007 AS-10 and documentation publicly available.¹³ No publication is permitted from the AS-10. However, it enables researchers to familiarize themselves with the AHDSS, determine its suitability for their work, and make a more-easily reviewed request to use the full AHDSS. The Agincourt review process is now much more efficient. In addition, the new extraction program enables the Agincourt team to create tailored full data sets expeditiously.

Intensive Training in Longitudinal Data Management and Analysis

Given the complexity of the structure of longitudinal data and techniques for longitudinal analysis, students need technical training to make appropriate use of the AS-10 or similar databases. CU researchers led the development and coordination of an intensive 3-week short course on longitudinal data management and analysis designed specifically to help students use AHDSS data. The course was given at CU in 2006 and 2007, with 16-19 students attending from four institutions with long-term research collaborations - Wits, CU, Brown University and the African Population and Health Research Center (APHRC, Nairobi, Kenya). It was taught by an inter-institutional faculty and included training in statistical analysis, data management, and use of the AS-10. In addition, students were exposed to international networks of students and faculty from various disciplines working on important global health issues. The course website, which includes readings, datasets, lecture notes, sample analysis programs, and other reference materials, remains available indefinitely, providing on-going research support to students.¹⁴

Wits, as part of its improvement of School of Public Health training facilities, installed a computer laboratory which accommodates 30 students. The course was offered there in 2008, taught jointly by faculty from CU, Wits, and APHRC. The course expanded to include 30 students and a second comparable sample database from the APHRC Nairobi Urban Health and Demographic Surveillance system prepared by APHRC and Agincourt team members. A key indicator that this approach is meeting student needs is the demand for the 2008 course: eligible students had to be turned away. This course has become a regular part of the Wits curriculum.

Resources

Resources for this project came from Wits and CU and grants to both institutions to improve access to the AHDSS and strengthen Wits training in population studies and public health. Creation of the annual AS-10 is now included in the yearly update of the AHDSS. Wits faculty participated in the course and will offer it in the future. The approach, therefore, is now sustainable without further external resources.

Conclusion

The need for increased training of public health professionals and increased research in Africa is clear. We have described a new model, developed at the University of the Witwatersrand with international collaboration, for linking research sites and training programs that involves preparation of user-friendly datasets accompanied by a short course in data management and analysis. The Wits model has demonstrated:

- (1) *Increased student and advanced researcher use of information from a longitudinal population study.* Students from several disciplines at Wits are now carrying out research using the AHDSS to complete masters and PhD level training.
- (2) *Acceleration of public health student training.* The course expanded student research opportunities and facilitated completion of their research. One course-evaluation reported:
"I had no idea how to start the analysis or how to incorporate it into my study. After the training course I now know how to analyze the data... which modules of the dataset would be applicable for my study, how to incorporate the data into my study, how to proceed with my analysis on a full sample dataset."
- (3) *Increased international and interdisciplinary research collaborations.* Faculty across the participating institutions became familiar with the research potential of the AHDSS and the comparable Nairobi system. This interest has resulted in collaborations which also generate grants that increase funding for research and student support.

Opportunities for replicating this research-ready dataset/ linked training model exist in many parts of the world. The INDEPTH network is today approaching some 40 field sites in Africa, Asia, Oceania and South America; all generate longitudinal, household-level demographic, health and socio-economic data and many are associated with training programs.⁶ Its leadership is actively encouraging replication of this model.

Box 1. Summary of Lessons Learned

- Longitudinal study sites can increase use of data, especially within public health training programs, by preparing user-friendly documentation and anonymized sample datasets on a regular basis (e.g., annually).
- Technical training targeted to these datasets is essential to encouraging their use by students and faculty.
- Strong international South-South and North-South research and training collaborations can facilitate the process of linking research and training.

Funding: Core funding for the MRC/Wits-Agincourt health and socio-demographic surveillance system is provided by The Wellcome Trust, UK (069683/Z/02/Z); The Andrew W. Mellon and William and Flora Hewlett Foundations, USA; the University of the Witwatersrand and Medical Research Council, South Africa; and the National Institute on Aging, NIH, USA. Funds for the support of training at Wits, the development of the training data, the delivery of the short course, and support for international research collaborations were provided to the University of Colorado by The William and Flora Hewlett Foundation (2005-6516 and 2006-8377), U.S. National Institute on Aging (1 R03 AG028640-01), The John E. Fogarty International Center (1 D43 TW007694-01), and The Andrew W. Mellon Foundation (20400663). Funding agreements have not limited the development of these programs or the research associated with them.

Competing Interests: None declared.

References

1. Sadana R, Petrakova A, Shaping public health education around the world to address health challenges in the coming decades. *Bulletin of the World Health Organization*. Dec 2007;85(12):902.
2. IJsselmuiden CB, Nchinda TC, Duale S, Tumwesigye NM, Serwadda D. Mapping Africa's advanced public health education capacity: the AfriHealth project. *Bulletin of the World Health Organization*. Dec 2007;85(12):914-922.
3. Petrakova A, Sadana R. Problems and progress in public health education. *Bulletin of the World Health Organization*. Dec 2007;85(12):963-964.
4. Voyi K. Producing a capable workforce. *Bulletin of the World Health Organization*. Dec 2007;85(12):967.
5. Haines A, Huttly S. The challenges of scaling-up. *Bulletin of the World Health Organization*. Dec 2007;85(12):967-968
6. INDEPTH: An International Network of field sites with continuous Demographic Evaluation of Populations and Their Health in developing countries [homepage on the Internet]. [updated 2007 Aug 27, cited 2007 Aug 27]. Available from: <http://www.indepth-network.org/default.asp>.
7. School of Public Health [homepage on the Internet]. Johannesburg: The University of the Witwatersrand; [cited 2009, Jan 24]. Available from: <http://web.wits.ac.za/Academic/Health/PublicHealth/>.
8. Agincourt Health and Population Unit [homepage on the Internet]. Johannesburg: The University of the Witwatersrand; [cited 2009, Jan 24]. Available from: <http://web.wits.ac.za/Academic/Health/PublicHealth/Agincourt/agincourthome.htm>
9. Institute of Behavioral Science [homepage on the Internet]. Boulder: The University of Colorado; [updated 2008 Oct 27, cited 2009 Jan 24]. Available from: <http://www.colorado.edu/ibs/>.

10. African Population Studies Research Program [homepage on the Internet]. Boulder: The University of Colorado Institute of Behavioral Science; [updated 2007 Feb 14, cited 2009 Jan 7]. Available from: www.colorado.edu/ibs/aps/.
11. Kahn K, Tollman SM, Collinson MA, Clark SJ, Twine R, Clark BD, Shabangu M, Gomez-Olive FX, Mokoena O, Garenne ML. Research into health, population and social transitions in rural South Africa: Data and Methods of the Agincourt Health and Demographic Surveillance System. *Scand J Pub Health*. 2007;35 Suppl 69:8-20.
12. Tollman SM, Kahn K, editors. *Scand J Pub Health*. 2007;35 Suppl 69:1-187.
13. Agincourt Health & Demographic Surveillance System [homepage on the Internet]. Johannesburg: The University of the Witwatersrand; [cited 2009 Jan 24]. Available from: <http://www.agincourt.co.za/DataSection/test/index.htm>
14. African Population Studies Research Program [Short Courses]. Boulder: The University of Colorado Institute of Behavioral Science; [updated 2008 June 3, cited 2009 Jan 24]. Available from: http://www.colorado.edu/ibs/aps/short_courses.html.