

# IBS

POPULATION PROGRAM ■  
INSTITUTE OF BEHAVIORAL SCIENCE ■  
UNIVERSITY OF COLORADO AT BOULDER ■



***WORKING PAPER***

## **Linking Research and Training through Research Data and Transnational Collaborations**

Jill R. Williams  
Enid J. Schatz  
Benjamin D. Clark  
Mark A. Collinson  
Samuel J. Clark  
Jane Menken  
Kathleen Kahn  
Stephen M. Tollman

September 2007

# Linking Research and Training through Research Data and Transnational Collaborations

Jill R. Williams,<sup>1,2</sup>  
Enid J. Schatz,<sup>a,b,3</sup>  
Benjamin D. Clark,<sup>b</sup>  
Mark A. Collinson,<sup>b</sup>  
Samuel J. Clark,<sup>a,b,4</sup>  
Jane Menken,<sup>a,b</sup>  
Kathleen Kahn,<sup>b</sup>  
Stephen M. Tollman<sup>b</sup>

## Corresponding author and contact information:

Jill R. Williams; UCB 483; Boulder, Colorado 80309-0483; 303-492-5253;  
Jill.Williams@Colorado.edu

---

<sup>1</sup> Population Program, University of Colorado-Boulder, UCB 483, Boulder, CO, 80309.

<sup>2</sup> MRC/Wits University Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, University of the Witwatersrand, 10th Floor, Room 10, 7 York Road, Parktown 2193 South Africa, Agincourt@sph.wits.ac.za

<sup>3</sup> University of Missouri-Columbia, School of Health Professions, 420 Lewis Hall, Columbia MO 65203

<sup>4</sup> University of Washington, Department of Sociology 202 Savery Hall, Box 353340, Seattle WA 98195-3340

## **Abstract**

A growing number of schools of public health are connected to Health and Demographic Surveillance System (HDSS) field sites in developing countries. These sites collect longitudinal health and demographic data, providing information of considerable use for health program and policy evaluation. Linking training programs with HDSS sites provides important opportunities - and challenges - for students to learn first-hand about health in the populations they intend to serve.

This case study describes efforts to overcome the barriers to linking research and training by making HDSS data more accessible and providing intensive training focused on using HDSS data. Specifically, we discuss the development of a 10%-sample training database from the Agincourt Health and Demographic Surveillance System, in rural northeast South Africa, run by the University of the Witwatersrand School of Public Health. Further, we outline the components of a concordant transnational intensive short course on longitudinal data analysis offered at the University of Colorado at Boulder in 2006 and 2007.

Early results from this collaborative effort demonstrate that this program has the potential to improve student research, link students to an international network of scholars, and increase student retention and matriculation. Importantly, the model developed is one that can be adapted to other partnerships between schools of public health and health and demographic research field sites.

## **Introduction**

A growing number of schools of public health are connected to Health and Demographic Surveillance System (HDSS) field sites in developing countries. These sites collect longitudinal health and demographic data at least annually, providing information of considerable use for health program and policy evaluation.<sup>1</sup> Linking training programs with HDSS sites provides important opportunities for students to learn first-hand about health in the populations they intend to serve. However, doing so poses challenges to both sites and their associated training programs. HDSS sites face challenges in protecting the confidentiality of human subjects and finding the time and resources to extract multiple usable research datasets with user-friendly documentation. Schools of Public Health face challenges in providing appropriate training for faculty and students to use complex HDSS data. As a consequence, HDSS data are often underutilized in public health teaching programs.

This case study describes efforts to address these challenges and work toward two mutually reinforcing goals:

- ◆ To improve public health training through development of an experiential short course that offers students the opportunity to learn about and carry out research using data from a health and demographic surveillance system; and
- ◆ To increase public health research by improving availability of and access to data of this type.

This effort is part of collaboration between the University of the Witwatersrand School of Public Health (Wits) and its MRC/Wits Rural Public Health and Health Transitions Research Unit, which runs the Agincourt Health and Demographic Surveillance System (AHDSS), and the

African Population Studies Research and Training Program (APS) at the University of Colorado at Boulder (CU). The program described below is a model that can be adapted by other research and training programs to facilitate community-based, problem-oriented teaching approaches addressing health challenges; steer student-led research toward global and national priorities; to add public health learning to other schools' curricula, and increase the international network of scholars working on developing country population health issues.

### **Wits/CU Training and Research Collaborations**

*Wits School of Public Health, the MRC/Wits Rural Public Health and Health Transitions Research Unit (known as the Agincourt Unit), and the Agincourt Health and Demographic Surveillance System*

The AHDSS was established in rural north-east South Africa in the early 1990s to provide reliable longitudinal population-based health and demographic information to inform health sector priorities and the restructuring of the health care system. A baseline census was conducted in 1992 with yearly follow-up to record vital events (births, deaths, in- and out-migrations). An important contribution of the AHDSS, given weak vital registration systems, is provision of reliable information on cause of death for health system planning. Notably, verbal autopsy methodology has been developed and verbal autopsies have been performed for all reported deaths. Mortality trends in the twenty-one villages of Agincourt have, therefore, been documented with some certainty and cause of death assigned to a significant portion of deaths in the area.<sup>2-5</sup>

The AHDSS is recognized as an important tool for studying the dynamics of health and socio-economic transition, health and health behavior, evaluating health and development programs, and conducting community trials in rural South and southern Africa. A 2007 supplement to the *Scandinavian Journal of Public Health* is devoted to research from the site.<sup>6</sup> Studies based on

AHDSS information include prevalence of stroke and its risk factors,<sup>7</sup> burden of pulmonary tuberculosis and associated health seeking behavior,<sup>8,9</sup> health care utilization, migration and mortality,<sup>10</sup> and social impacts of HIV/AIDS.<sup>11-13</sup> Since 1997, the annual census update has included special modules that regularly track important social and economic variables at the household and individual level. These modules expand the range and depth of the ADHSS data, allowing for broader range of health and demographic research. This expansion increases the value and utility of the data as well as its complexity, and therefore also increases the burden of data management on the Agincourt Unit.

To help increase local technical and research capacity, the Agincourt Unit has played a leading role in efforts to create postgraduate degree programs in the measurement sciences at Wits. These programs include the MSc in Population-based Field Epidemiology in the School of Public Health (offered in partnership with the INDEPTH Network<sup>14</sup>), and the Honours-MA and PhD in Demography and Population Studies located in the School of Social Sciences. Agincourt and the AHDSS serve as an important training resource for both programs, with students nested in the AHDSS or related projects. Tollman, director of the Agincourt Unit, also sought international research and training collaborations that would contribute to the training programs at Wits and research in the Unit. Collaboration with the University of Colorado African Population Studies Research and Training Program is one result of this outreach.

#### *CU African Population Studies Research and Training Program (APS)*

APS, established in 2001 within the Population Program of the Institute of Behavioral Science, was designed to address three key needs: 1) the shortage of population experts in sub-Saharan Africa trained to analyze important social and health issues affecting the continent, 2) the lack of high-quality training programs within the region, and 3) the need for increased research and teaching on African health and population issues within U.S. universities.<sup>15</sup> The program has

helped support research in Agincourt and development of training programs at Wits. APS post-doctoral fellows have provided administrative assistance, taught courses in both the School of Public Health and School of Social Sciences, and contributed to data management and research in Agincourt.

To further research on population and health in sub-Saharan Africa, Wits and CU joined with the African Population and Health Research Center (APHRC), located in Nairobi, Kenya, and Brown University to establish a Colloquium on Emerging Population Issues, held annually since 2003. Common themes in discussions at the colloquium were the richness of the data being collected by Wits and APHRC, the difficulty of gaining access to and analyzing them, and the potential training opportunities the data provide. African students and scholars repeatedly drew attention to the limited availability of quality training in African settings that also provides opportunity to link effectively with international networks, ideas, and debates. Some of the training and data challenges are discussed below, followed by a description of our collaborative efforts to overcome them.

### **Access to AHDSS Data**

Currently a researcher or student submits a request for part of the AHDSS data; the Agincourt team reviews the proposal and determines whether it fits into the overall research agenda; and if the proposal is approved, an Agincourt staff member writes a unique script to extract the a fully anonymized dataset. There are several challenges in this procedure. First, obtaining approval is often delayed because the Agincourt team is fully occupied with a research agenda that includes maintaining the quality of the AHDSS. Second, those requesting data rarely have experience using AHDSS data; their limited knowledge of the database's complexity often leads to underspecified data requests. Difficulty in deciphering a request adds more time to the process. As a consequence researchers feel neglected and the Agincourt team feels constant pressure to produce datasets. Finally, no standard procedure is in place for ensuring that researchers outside

the Agincourt Unit meet its standards for protection of human subjects, leaving Agincourt scientists reluctant to release data except to direct collaborators.

In 2004, the authors proposed working toward a more regulated system that would allow broader access to the database and reduce the demand for the small Agincourt staff to draw individual datasets. This system would have a basic script that makes extraction of sub-sample and full datasets (stripped of individual identifiers) simpler and more automated. The eventual plan is to have an anonymized 10%-sample of AHDSS data with full documentation on a password-protected website. Researchers who agree to use the data for exploratory purposes only would be given a password. They could then explore the data and be better able to specify the dataset they need. Following review of the data request, the Agincourt team would then extract a complete anonymized dataset for the research project.

The Agincourt training database described below is an important first step in meeting the aim of wider access to longitudinal data. It is expected that increased access will lead to increased research output and ultimately to a more complete understanding of the lives of rural South Africans that will inform public health programs and policies. For students, timely access to data for research projects will likely decrease the length of their programs and thereby increase student productivity, retention and matriculation.

### **Data Development and Training Challenges**

The AHDSS training database is designed to mimic the full AHDSS database. It contains the same tables and relationships between tables, maintaining the structure and integrity of the full database and includes a 10%-sample of the units of observation in the AHDSS. The unit of observation is a geographic location where individuals and social groups reside. The sample is stratified to include 10% of locations in each of the site's twenty-one villages. It is validated by

comparing the household mean number of deaths, pregnancies, in-migrations, out-migrations, and household size against the entire database. If any sample mean falls outside one standard deviation from the full database mean, the sample is discarded and a new one drawn. The process is limited to 1000 draws, at which point the final sample is used. This validation process is intended to ensure that rates using event counts are comparable between the training and full AHDSS databases.

Some variables that are easily misinterpreted or too sensitive in nature for a public dataset are excluded from the training dataset. The sample is anonymized by replacing the original ten-digit personal identifiers with five-character identifiers generated using base-23 numbers. Concurrent to development of the dataset was construction of user-friendly documentation. The final result is an anonymized 10%-sample that retains the relational, temporal and data integrity characteristics of the full database, which can be used for training and exploratory research purposes.

Increasing availability of data does not ensure that students and researchers will be able to use them. In particular, the structure of longitudinal data is complex, as are the techniques used for longitudinal analysis. In fact, although Wits School of Public Health faculty wanted datasets their students could use to address questions specified by Agincourt scientists, they felt insufficiently prepared to support students working with this complex database. What was clearly needed was a way for faculty and graduate students to acquire knowledge of the Agincourt data and the technical skills necessary to use them.

### **Intensive Training in Longitudinal Data Management and Analysis**

In 2006 and 2007 the CU APS offered an intensive 3-week course to students from Wits, CU, Brown University and AHPRC intended to (1) provide training in longitudinal data management and analysis through use of the AHDSS and (2) strengthen ties among the participant institutions.

The course, coordinated by APS post-doctoral fellows and delivered by CU and Wits faculty, had the following components.

*Substantive introduction to rural population issues using the Agincourt, South Africa rural sub-district as a primary reference:* A particular goal of the course was to create a new cohort of students interested in research on African population and health issues. This component included readings, lectures, and CU/Wits faculty research presentations on issues such as HIV testing, AIDS and use of the environment, and pensions and the role of older women. These research presentations provided important examples of research designs that use HDSS data for a sampling framework.

*Training in statistical analysis, data management, and use of the AHDSS Training Dataset:* Participants gained hands-on experience using longitudinal methods to analyze health and demographic trends and their determinants. AHDSS training data were used in lectures and guided exercises to introduce students to statistical analysis using STATA and train them in managing longitudinal data using STATA and Microsoft Access. Techniques covered included descriptive analysis of fertility and mortality trends, event history analysis, and hazard modeling.

*Academic skills training:* Sessions on developing research questions, conducting literature reviews, writing-up research results, and giving presentations were held. In the second week, students developed their own research questions, constructed data sets, and conducted preliminary analysis. Collaborative learning and research was encouraged throughout the course and each group gave a brief presentation on its research project. In the third (non-resident) week of the course, participants completed a 7-10 page research paper describing their topic, analysis, and results. Students left the course having completed a simulation of the

research process – learning about a cultural setting, learning about a data set, formulating research questions, extracting the necessary data, conducting analyses, and presenting findings in oral and written form.

*Professional networking:* The course focus on collaboration provided participants a unique opportunity to expand their professional networks. Fifteen faculty members, including one each year from Wits, served as consulting faculty and met with students throughout the course to offer advice on current and future research. Research presentations, meetings with consulting faculty, and informal social events facilitated networking among students and junior and senior faculty.

*On-line Resources:* A password-protected class website was created that includes readings, datasets, lecture notes, sample STATA programs, and other reference materials. The website remains available indefinitely, thereby providing on-going research support to students.

The 2006 and 2007 participants now serve as local resources in their home institutions for others planning to use the AHDSS data. In 2008 the course will be offered at Wits with support from CU faculty. Agincourt and APHRC staffs are preparing a training database from the Nairobi Urban Health and Demographic Surveillance system, run by APHRC, which will be comparable to the AHDSS training database. Thus, African data from both rural and urban (slum) settings will be available to 2008 course participants.

## **Conclusion**

Anticipated outcomes from the course include:

- ◆ Better utilization of important health and demographic data
- ◆ Facilitation of interdisciplinary work

- ◆ Improved technical training for graduate students
- ◆ Embedding of students in an expanding international network of scientists – thereby increasing retention, early completion, collaborations, and publications
- ◆ Increased admissions and increased quality of students attracted by opportunities for a new kind of graduate education.

Early results from this program are encouraging. In 2006, an APHRC participant decided to apply to the SPH at Wits after learning about the AHDSS through the short course. She was awarded a competitive three-year full fellowship to work toward her PhD. Her fieldwork, conducted in Agincourt, was completed within one year. In 2007, Wits students from four departments (Public Health, Maths, Geography, and Demography) were accepted in the short course based in part on their interest in using the AHDSS in their own research projects. By familiarizing students with Agincourt, the AHDSS and published research in the field, the short course adds public health curricula to other departments.

Students from Wits also gained an important introduction to the AHDSS that will facilitate completion of their research. One student wrote:

“I had no idea how to start with the analysis or how to incorporate it into my study. After the training course I now know how to analyze the data using both MS Access and STATA, 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.”

These students gained greater capacity to manage longitudinal data, which relieves pressure on the Agincourt data manager and helps facilitate successful student-led research.

Students are also guided, in the substantive portion of the course, toward global and national health priorities. Students learning about African health issues for the first time are attracted to the field. Five CU students attended the 2007 short course; none had worked on African health issues. Two students began collaborations with faculty working with Agincourt data and a third is planning qualitative research in Agincourt. As one student from CU wrote, the course influenced her interests and likely her future career path:

“...what this course reminded me of was why I chose [graduate school] in the first place—to do research that can be used to make a difference. Learning about the Agincourt HDSS and the ways in which research is undertaken and utilized by the community resounded with my approach to sociology and to the world....This course has redefined my graduate school path and definitely redefined it for the better.”

Students have established their own on-line discussion group to maintain their network as they progress through graduate school and beyond, suggesting that the course is an effective tool for increasing peer networks. Further evaluation of this approach will be made as students are followed in their careers. This model, which breaks away from traditional courses by presenting students with the full research process and fresh data, seems likely to make a difference – and it is replicable using data from other health and demographic surveillance systems and in other schools of public health.

The renewed international focus on improving health information systems through initiatives such as the Health Metrics Network will likely increase demand for health professionals trained in measurement sciences.<sup>16</sup> Existing data, therefore, are an important resource for public health training programs. The INDEPTH network is today approaching some 40 field sites in Africa, Asia, Oceania, and South America, all of which generate longitudinal, household level

demographic, health and socio-economic data on an ongoing basis. These studies can provide an exceptional data source for students in public health institutions worldwide, making it worthwhile to consider how best to render such data sufficiently usable and accessible to students. The case-study described here provides one promising example that can be adapted to other contexts. In doing so, academic public health leaders will make tangible contributions to the application of such data to local and global health priorities and, in the process, train students to use longitudinal data for research and development, program and policy evaluation and – for that matter – the monitoring of progress towards the Millennium Development Goals.

**Funding:** Core funding for the AHDSS is provided by The Wellcome Trust, UK (058893/Z/99/A) with additional funding for research and training at Wits from the National Institute on Aging of the NIH, USA; The Andrew W. Mellon Foundation, the William and Flora Hewlett Foundation, the University of the Witwatersrand, and Medical Research Council SA. 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. Sankoh O, Kahn K, Kiwanuka N, Mwageni E, Ngom P, Nyarko P, Simba D, editors. *Population and Health in Developing Countries Volume 1: Population, Health, and Survival at INDEPTH sites*. Ottawa: IDRC; 2002.
2. Kahn K, Tollman SM, Garenne M, Gear JSS. Who dies from what? Determining cause of death in South Africa's rural northeast. *Trop Med and Int Health*. 1999;4:433-441.
3. Kahn K, Tollman SM, Garenne M, Gear JSS. Validation and application of verbal autopsies in a rural area of South Africa. *Trop Med and Int Health*. 2000;5(11):824-831.
4. Kahn K, *Dying to Make a Fresh Start: Mortality and Health Transition in a New South Africa* [dissertation]. Umeå (Sweden): Umeå University; 2006.
5. Kahn K, Garenne ML, Collinson MA, Tollman SM. Mortality trends in a new South Africa: Hard to make a fresh start. *Scand J Pub Health*. 2007;35 Suppl 69:26-34
6. Tollman SM, Kahn K, editors. *Scand J Pub Health*. 2007;35 Suppl 69:1403-4948.
7. The SASPI Team. The prevalence of stroke survivors in rural South Africa: results from the Southern Africa Stroke Prevention Initiative (SASPI) Agincourt field site. *Stroke*. 2004;627-632
8. Pronyk PM, Makhubele MB, Hargreaves JR, Tollman SM, Hausler HP. Assessing health seeking behaviour among tuberculosis patients in rural South Africa. *Int J Tuberc Lung Dis*. 2001;5(7):619-627
9. Pronyk P, Kahn K, Hargreaves J, Tollman S, Collinson M, Hausler HP, Porter JDH. Undiagnosed pulmonary tuberculosis deaths in rural South Africa. *Intl J Tuberc Lung Dis*. 2004; 8(6):796-799
10. Clark SJ, Collinson MA, Kahn K, Drullinger K, Tollman SM. Returning home to die: Circular labour migration and mortality in South Africa. *Scand J Pub Health*. 2007;35 Suppl 69: 35-44
11. Hunter LM, Twine W, Patterson L. 'Locusts are now our beef:' Adult mortality and household dietary use of local environmental resources in rural South Africa. *Scand J Pub Health*. 2007;35 Suppl 69:165-174
12. Posel D, Kahn K, Walker E. Living with death in a time of AIDS: A rural South African case study. *Scand J Pub Health*. 2007;35 Suppl 69:138-146
13. Schatz EJ, Ogunmefun C. Caring and Contributing: The role of older women in multi-generational households in the HIV/AIDS era. *World Dev*. 2007;35(8):1390-1403.
14. 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>.
15. African Population Studies Research Program [homepage on the Internet]. Boulder: The University of Colorado Institute of Behavioral Science; [updated 2007 Feb 14, cited 2007 Aug 22]. Available from: [www.colorado.edu/ibs/aps/](http://www.colorado.edu/ibs/aps/).

16. Health Metrics Network [homepage on the Internet]. World Health Organization c2007 [cited 2007 Aug 30]. Available at: <http://www.who.int/healthmetrics/en/>.