CO/\CTION
PUBLISHING

Improving public health training and research capacity in Africa: a replicable model for linking training to health and socio-demographic surveillance data

Jill R. Williams[1,2]*, Enid J. Schatz[1,2,3], Benjamin D. Clark[2,4,5], Mark A. Collinson[2,6], Samuel J. Clark[1,2,7], Jane Menken[1,2], Kathleen Kahn[1,2,6] and Stephen M. Tollman[1,2,6]

[1]Population Program, Institute of Behavioral Science, University of Colorado-Boulder, Boulder, CO, USA; [2]MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; [3]School of Health Professions, University of Missouri, Columbia, MO, USA; [4]London School of Hygiene and Tropical Medicine, University of London, London, UK; [5]TAZAMA Project, National Institute of Medical Research, Tanzania; [6]Department of Public Health and Clinical Medicine, Umeå Centre for Global Health Research, Umeå University, Umeå, Sweden; [7]Department of Sociology, University of Washington, Seattle, WA, USA

Background: Research training for public health professionals is key to the future of public health and policy in Africa. A growing number of schools of public health are connected to health and socio-demographic surveillance system field sites in developing countries, in Africa and Asia in particular. Linking training programs with these sites provides important opportunities to improve training, build local research capacity, foreground local health priorities, and increase the relevance of research to local health policy.
Objective: To increase research training capacity in public health programs by providing targeted training to students and increasing the accessibility of existing data.
Design: This report is a case study of an approach to linking public health research and training at the University of the Witwatersrand. We discuss the development of a sample training database from the Agincourt Health and Socio-demographic Surveillance System in South Africa and outline a concordant transnational intensive short course on longitudinal data analysis offered by the University of the Witwatersrand and the University of Colorado-Boulder. This case study highlights ways common barriers to linking research and training can be overcome.
Results and Conclusions: This collaborative effort demonstrates that linking training to ongoing data collection can improve student research, accelerate student training, and connect students to an international network of scholars. Importantly, the approach can be adapted to other partnerships between schools of public health and longitudinal research sites.

Keywords: *Africa; public health training; data accessibility; research capacity; health policy*

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 health professionals are trained to design, implement, and evaluate public health policies and programs or to contribute to the 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). Additionally, local research capacity development has been identified as an

important step in ensuring that research is relevant to local community health concerns and that local policy makers make use of this evidence (6).

Providing high quality, current, and relevant data with which trainees can work is an important challenge. However, a growing number of schools of public health are connected to field sites in developing countries that carry out health and socio-demographic surveillance, collecting data at least yearly (7). This information is being used to inform health policy, to ensure that services are responsive and accountable to local needs, and to help address health inequities (8–10). Trainee access to field sites and data can provide exceptional opportunities for students to learn firsthand about health in the populations they intend to serve.

Connecting training programs to demographic surveillance sites poses challenges to both partners. Field sites may lack resources to support training (including involving students in field research and/or preparing data for student use) and must ensure that trainee datasets protect confidentiality of human subjects. Training programs need to provide students and supervisors with the resources and technical skills to analyze the complex longitudinal data produced by demographic surveillance sites. This paper describes our experience overcoming these challenges while connecting demographic and public health graduate training at the University of the Witwatersrand (Wits), South Africa, to the Agincourt Health and Socio-demographic Surveillance System (Agincourt HDSS). This effort is part of research and training collaborations between the Wits School of Public Health and its MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt; The MRC/Wits-Agincourt Unit), and the University of Colorado–Boulder (CU) Institute of Behavioral Science, United States, and its African Population Studies Research and Training Program (11–14).

Wits school of public health and the Agincourt health and socio-demographic surveillance system

The Agincourt HDSS was established by the Wits School of Public Health in rural north-east South Africa in the early 1990s with the goal of supporting and documenting the restructuring of the health care system in post-apartheid South Africa (15). To this end, evidence on clinic catchment populations and prevailing health problems was required to inform priority-setting and program planning. To do so, reliable population-based health and demographic information was essential, especially since vital registration systems in such rural settings were weak. The Agincourt HDSS began with a baseline census in 1992 and continues with annual updates of household membership and vital events (births, deaths, in- and out-migrations). It currently

covers a subdistrict of 25 villages, with some 14,800 households and a population of over 84,000 people (15).

An important component of the Agincourt HDSS is reliable information on cause of death. A verbal autopsy (VA) has been conducted on all reported deaths in the subdistrict since 1992 and a cause has been assigned to the majority of deaths (15–18). Mortality trends in the Agincourt setting have, therefore, been documented with some certainty, demonstrating a significant simultaneous rise in chronic infectious illness (HIV/AIDS and TB), and non-communicable disease (vascular conditions and associated risk factors in particular) (15–20). Since 2000, the census has included special modules that regularly collect important household and individual health, social, and economic data (21). These modules, samples of which are listed in Table 1, expand the range and depth of the Agincourt HDSS data and increase their potential application to health policy and program development. They also increase the complexity of the longitudinal data available to researchers.

The Agincourt HDSS, therefore, provides information on the breadth and depth necessary for understanding health and health behavior, monitoring health and population change, studying dynamics of health and socioeconomic transition, preparing and conducting community trials, and evaluating health and development programs in the region (12, 20, 21). Recent studies include prevalence of stroke and its risk factors (22); burden of pulmonary tuberculosis and associated health seeking behavior (23, 24); health care utilization, migration, and mortality (25, 26); and the social impact of HIV/AIDS (27–29).

In recognition of its work, the MRC/Wits-Agincourt Unit was designated a South African Medical Research Council (MRC) Unit in 2003, formally termed the MRC/

Table 1. A sample of special survey modules collected with the Agincourt census update[a]

Special survey module	Years collected
Household assets	2001, 2003, 2005, 2007, 2009
Child health care utilization and morbidity	2006
Child social grant access	2002, 2005, 2008
Father support status (extended)	2007
Food security	2004, 2007, 2010
Health care utilization	2003, 2006, 2010
Labor participation	2000, 2004, 2008
Physical and cognitive functioning (adult)	2006, 2010

[a]For more information, see http://www.agincourt.co.za/Data Section/DataCollection.htm.

Wits Rural Public Health and Health Transitions Research Unit. This connection assists efforts to have findings from the Agincourt HDSS and associated projects inform health program and policy development, and helps ensure that policy needs of the South African government are considered when shaping the MRC/Wits-Agincourt Unit research agenda.

To help increase local and regional technical and research capacity, the Wits School of Public Health and its MRC/Wits-Agincourt Unit played a leading role in creating postgraduate degree programs in measurement sciences. These include an MSc in Population-based Field Epidemiology (offered in partnership with the INDEPTH Network) and an Honors-MA and PhD program in Demography and Population Studies (offered in conjunction with the School of Social Sciences). These interdisciplinary programs were set up to expand the cadre of public health professionals capable of conducting interdisciplinary population health research drawing on a mix of measurement sciences and social research methods. The MRC/Wits-Agincourt Unit serves as an important training resource for both programs. Masters level students work with Wits faculty on research using Agincourt HDSS data while PhD students are nested within a range of ongoing field-based research in the Agincourt site. Students benefit greatly from the MRC/Wits-Agincourt Unit research infrastructure, which provides access to research facilities, trained fieldworkers, and sampling frames.

The Agincourt HDSS provides invaluable opportunities for training new PhD researchers in the population and health science programs at Wits, as well as Masters level (MSc, MPH, MA) students. However, initial efforts to involve students in research using Agincourt HDSS data revealed obstacles to gaining access to and then analyzing information from this complex data system. Important barriers are discussed below, followed by a description of our efforts to overcome them.

Access to Agincourt HDSS data

Prior to 2005, the steps for obtaining Agincourt HDSS data were:

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

A significant challenge was that students and supervisors lacked familiarity with longitudinal data generally and Agincourt HDSS data specifically. Underspecified requests were submitted, which the MRC/Wits-Agincourt staff found difficult to decipher. The result was often a long process where data requests were revised and resubmitted and datasets often had to be reconfigured many times to suit a research project. Consequently, time to approval and data preparation was often 6 months or more, making it difficult for students to complete research within the time limits of their academic programs. A further challenge was the burden dataset preparation placed on the MRC/Wits-Agincourt staff. These problems led to underutilization of the Agincourt HDSS by students and even experienced researchers.

In 2004, the authors collectively began reformulating the process of obtaining Agincourt HDSS data to include both technical training and the possibility of exploratory analysis of the data prior to requesting a custom dataset. A crucial first step was the 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 Agincourt HDSS with anonymized individual identifiers to protect human subject confidentiality. The sample database maintains the structure of the full relational Agincourt HDSS database. Its means fall within one standard deviation of Agincourt HDSS means for key variables; thus, results are expected to be comparable between sample and full Agincourt HDSS 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.

Importantly, the AS-10 and documentation are now publicly available and easily accessible on the Internet (30). No publication is permitted from the AS-10. However, it enables researchers to familiarize themselves with the Agincourt HDSS, determine its suitability for their work, and make a more detailed and appropriate request to use the full Agincourt HDSS. These changes have made the Agincourt review process much more efficient, since the need to repeatedly revise requests and datasets has been greatly reduced. In addition, the new extraction program enables the Agincourt team to create custom full datasets expeditiously.

The process of developing datasets has also been facilitated by hiring a data specialist whose specific function is to extract data for students. The salary of the additional data specialist is paid from a fund that supports student fellowships. With the development of the Web site, database documentation, the AS-10, and the hiring of a new data specialist, the review process has halved from 6 to 3 months per request. Since the AS-10 is

now downloadable from the Agincourt data Web site (30), it is immediately accessible to all interested users. Students can begin writing programs and run sample analyses while awaiting their tailor-made dataset. Currently, about six students per year work on datasets tailor-made from the full Agincourt HDSS database.

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. The 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 AS-10 data. The course was given at CU in 2006 and 2007, with 16–19 students attending each year 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 repeated in 2008 at Wits where it was taught by an interinstitutional faculty to 30 students from the four institutions. Participants from Wits, CU, and Brown University were generally students enrolled in Masters or PhD programs. Attendees from APHRC were generally researchers with master's level training, not currently enrolled in a training program.

Each year participants gained hands-on experience using longitudinal methods to analyze health and demographic trends and their determinants. The AS-10 was used in lectures, guided exercises, and research projects 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. In 2008 a second comparable sample database from the APHRC Nairobi Urban HDSS was added to the course. The course Web site, which includes readings, datasets, lecture notes, sample analysis programs, and other reference materials remains available indefinitely, providing on-going research support to students (31).

In addition to technical training in longitudinal data management, a particular goal of the course was to create a cohort of students interested in research on African population and health issues and connect them to an international network of researchers engaged in this area. Therefore, the curriculum included an introduction to important health and demographic issues in Africa. This component included readings, lectures, and faculty research presentations on issues such as HIV testing, AIDS, migration, the environment, and aging. These research presentations provided important examples of research designs that both utilize HDSS data for a sampling framework and directly address important local health issues.

Since one of our goals is to create a cohort of students trained to do research on population and health in Africa, we followed up with the participants in the three short courses in 2006, 2007, and 2008. As shown in Table 2, as of 2010, we found that a large majority of participants have either completed advanced degrees, are continuing to work on their degrees, or have entered new advanced degree programs. Only four students who were enrolled in the advanced degree programs at the time they took the short course dropped out of those programs. And, of the 14 short course participants who were not already enrolled in advanced degree programs, 10 have since enrolled in PhD programs. The short course has clearly served students who are committed to pursuing and finishing advanced degrees in population and health sciences and these students reflect a growing capacity for health research in Africa.

Students themselves credit the short course for helping them complete research projects using longitudinal data including the Agincourt HDSS. According to one course evaluation, '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.'

It is important to note that initial resources for this project came from both Wits and CU, including grants to both institutions to improve access to the Agincourt HDSS and strengthen Wits training in population studies and public health. Creation of the annual AS-10 is now included as part of the yearly Agincourt HDSS update. Wits faculty participated in the short course and now offer it as part of their population and health curriculum. This approach to training, therefore, is now sustainable at Wits without further external resources.

Table 2. Short course participant follow-up, 2010

	Participants	Enrolled in MA/MSc/PhD at time of short course	Dropped program	Completed program	Continuing program	Has since entered PhD program
African	40	26	4	12	10	9
International	15	14	0	5	9	1
Total	55	41	4	19	19	10

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. This model simultaneously increases the availability of important HDSS data and builds local research capacity, which helps foreground local health priorities and increases the relevance of research to local health policy. The Wits model has also demonstrated:

1) *Increased researcher use of information from a longitudinal population study.* An expanded group of students and faculty from several disciplines at Wits and CU are now carrying out research using the Agincourt HDSS 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 training programs.
3) *Increased international and interdisciplinary research collaborations.* Faculty across the participating institutions became familiar with the research potential of the Agincourt HDSS and the comparable Nairobi system. This interest has resulted in further research and training collaborations, which have been funded by grants from agencies such as the National Institutes of Health, Fogarty International Center, The William and Flora Hewlett Foundation (USA), and The Wellcome Trust (UK).

Opportunities for replicating this research-ready dataset and linked training model exist in many parts of the world. The INDEPTH network at large is today approaching some 40 field sites in Africa, Asia, Oceania, and South America; all generate longitudinal, household-level demographic, health and socioeconomic data, and many are associated with training programs (7). Its leadership is actively encouraging replication of this model through its 1:10 Training initiative to make DSS data available on the web for pilot projects (7). The MRC/Wits-Agincourt Unit is also a member of the INDEPTH iSHARE network, whose goals include the development of new methodologies to share data effectively and efficiently. The iSHARE INDEPTH network includes six HDSS sites in Asia and Africa working to expand data sharing and increase comparability of data across contexts (32).

More work is needed, however, to ensure that wider access to data and increases in both local research and student training capacity translate into the development of evidence-based health policy. Students and trainees are undertaking important health and demographic research (33), but need further training to connect research with policy. To this end we are currently planning future collaborative training programs to further support the translation of research to policy makers (Box 1).

Box 1. Summary points

- Southern research capacity building is an important step in supporting evidence-based health policy formation.
- 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 encourage 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.

Conflict of interest and funding

References

1. Sadana R, Petrakova A. Shaping public health education around the world to address health challenges in the coming decades. Bull World Health Organ 2007; 85: 902.
2. IJsselmuiden CB, Nchinda TC, Duale S, Tumwesigye NM, Serwadda D. Mapping Africa's advanced public health education capacity: the AfriHealth project. Bull World Health Organ 2007; 85: 914–22.
3. Petrakova A, Sadana R. Problems and progress in public health education. Bull World Health Organ 2007; 85: 963–4.

4. Voyi K. Producing a capable workforce. Bull World Health Organ 2007; 85: 967.

5. Haines A, Huttly S. The challenges of scaling-up. Bull World Health Organ 2007; 85: 967–8.

6. Vincent R. Communicating health research: how should evidence affect policy and practice? Stronger links between researchers, policy makers and practitioners and increased southern research capacity are the keys to relevant evidence being taken up. Exchange Findings 5; 2009. Available from: http://www.healthcomms.org/pdf/findings5.pdf [cited 10 May 2010].

7. INDEPTH: An International Network of field sites with continuous Demographic Evaluation of Populations and Their Health in developing countries [homepage on the Internet]; 2010. Available from: http://www.indepth-network.org [cited 10 May 2010].

8. Tollman SM, Zwi AB. Health system reform and the role of field sites based upon demographic and health surveillance. Bull World Health Organ 2000; 78: 125–34.

9. Ngom P, Binka FN, Phillips JF, Pence B, Macleod B. Demographic surveillance and health equity in sub-Saharan Africa. Health Policy Plann 2001; 16: 337–44.

10. Chandramohan D, Shibuya K, Setel P, Cairncross S, Lopez AD, Murray CJL, et al. Should data from demographic surveillance systems be made more widely available to researchers? PLoS Med 2008; 5: 0169–0173.

11. School of Public Health, The University of the Witwatersrand [homepage on the Internet]; 2009. Available from: http://web.wits.ac.za/Academic/Health/PublicHealth/ [cited 10 May 2010].

12. Agincourt Health and Population Unit, The University of the Witwatersrand [homepage on the Internet]; 2009. Available from: http://web.wits.ac.za/Academic/Health/PublicHealth/Agincourt/agincourthome.htm [cited 10 May 2010].

13. Institute of Behavioral Science, The University of Colorado, Boulder [homepage on the Internet]; 2010. Available from: http://www.colorado.edu/ibs/ [cited 10 May 2010].

14. African Population Studies Research Program, The University of Colorado, Boulder [homepage on the Internet]; 2010. Available from: http://www.colorado.edu/ibs/aps/ [cited 10 May 2010].

15. Kahn K, Tollman SM, Collinson MA, Clark SJ, Twine R, Clark BD, et al. 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: 8–20.

16. Kahn K, Tollman SM, Garenne M, Gear JSS. Who dies from what? Determining cause of death in South Africa's rural northeast. Trop Med Int Health 1999; 4: 433–41.

17. 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: 824–31.

18. Kahn K. Dying to make a fresh start: mortality and health transition in a New South Africa. Dissertation, Umeå University, Umeå, Sweden, 2006.

19. 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: 26–34.

20. Tollman SM, Kahn K, editors. Health, population & social transitions in rural South Africa. Scand J Pub Health 2007; 35: 1–187.

21. Agincourt Health & Demographic Surveillance System, The University of the Witwatersrand [homepage on the Internet]. Available from: http://www.agincourt.co.za/DataSection/ [cited 10 May 2010].

22. 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: 35: 627–32.

23. 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: 619–27.

24. Pronyk P, Kahn K, Hargreaves J, Tollman S, Collinson M, Hausler HP, et al. Undiagnosed pulmonary tuberculosis deaths in rural South Africa. Intl J Tuberc Lung Dis 2004; 8: 796–9.

25. 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: 35–44.

26. Collinson MA, Wolff B, Tollman SM, Kahn K. Trends in internal labour migration from rural Limpopo Province, male risk behaviour, and implications for the spread of HIV/AIDS in rural South Africa. J Ethnic Migration Stud 2006; 32: 633–48.

27. 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: 165–74.

28. 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: 138–46.

29. 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: 1390–403.

30. Agincourt Health & Demographic Surveillance System [data downloads]. Available from: http://www.agincourt.co.za/DataSection/DataDownloads.htm [cited 10 May 2010].

31. African Population Studies Research Program [short courses]; 2009. Available from: http://www.colorado.edu/ibs/aps/short_courses.html [cited 10 May 2010].

32. INDEPTH Network iSHARE; 2009. Available from: http://www.indepth-ishare.org/ [cited 30 July 2010].

33. Agincourt Student Projects; 2009. Available from: http://web.wits.ac.za/Academic/Health/PublicHealth/Agincourt/Students.htm [cited 27 July 2010].

*Jill R. Williams
Population Program
UCB 483, Institute of Behavioral Science
University of Colorado-Boulder
Boulder, CO 80309-0483, USA
Tel: +1 303 492 5253
Email: Jill.Williams@Colorado.edu