



## Practice of Epidemiology

# Development and Validation of a Global Positioning System–based “Map Book” System for Categorizing Cluster Residency Status of Community Members Living in High-Density Urban Slums in Blantyre, Malawi

Peter MacPherson\*, Augustine T. Choko, Emily L. Webb, Deus Thindwa, S. Bertel Squire, Rodrick Sambakunsi, Joep J. van Oosterhout, Treza Chunda, Kondwani Chavula, Simon D. Makombe, David G. Lalloo, and Elizabeth L. Corbett

\* Correspondence to Dr. Peter MacPherson, Malawi-Liverpool-Wellcome Trust Clinical Research Programme, P.O. Box 30096, Chichiri 3, Blantyre, Malawi (e-mail: [p.macpherson@liverpool.ac.uk](mailto:p.macpherson@liverpool.ac.uk)).

Initially submitted May 5, 2012; accepted for publication September 4, 2012.

A significant methodological challenge in implementing community-based cluster-randomized trials is how to accurately categorize cluster residency when data are collected at a site distant from households. This study set out to validate a map book system for use in urban slums with no municipal address systems, where classification has been shown to be inaccurate when address descriptions were used. Between April and July 2011, 28 noncontiguous clusters were demarcated in Blantyre, Malawi. In December 2011, antiretroviral therapy initiators were asked to identify themselves as cluster residents (yes/no and which cluster) by using map books. A random sample of antiretroviral therapy initiators was used to validate map book categorization against Global Positioning System coordinates taken from participants' households. Of the 202 antiretroviral therapy initiators, 48 (23.8%) were categorized with the map book system as in-cluster residents and 147 (72.8%) as out-of-cluster residents, and 7 (3.4%) were unsure. Agreement between map books and the Global Positioning System was 100% in the 20 adults selected for validation and was 95.0% ( $\kappa = 0.96$ , 95% confidence interval: 0.84, 1.00) in an additional 20 in-cluster residents (overall  $\kappa = 0.97$ , 95% confidence interval: 0.90, 1.00). With map books, cluster residents were classified rapidly and accurately. If validated elsewhere, this approach could be of widespread value in that it would enable accurate categorization without home visits.

Africa; antiretroviral therapy; cluster-randomized trials; community-based studies; Global Positioning System; human immunodeficiency virus; maps

Abbreviations: ART, antiretroviral therapy; CI, confidence interval; CONDA-YAPA, Convenience in Delivery of ART: Yambirani Pakhomo study; CRT, cluster-randomized trial; GPS, Global Positioning System; Hit-TB, Hit-TB Hard Study.

Over the past 2 decades, there has been a large increase in the number of cluster-randomized trials (CRTs) undertaken (1–3). CRTs offer particular benefits to public health researchers (4), including the opportunity to measure the population impact of interventions (5), convenience in applying interventions to whole groups of persons (4), and the avoidance of contamination that might occur if individuals within populations were allocated randomly to receive an intervention (6).

Rapid urbanization in low- and middle-income countries has resulted in the proliferation of densely populated slums

that lack basic urban planning, including address systems. The United Nations Human Settlements Programme defines a slum household as lacking more than one of the following: adequate access to water and sanitation, sufficient living space, durability of housing, and security of tenure (7). Slums are important sites for infectious disease transmission and frequently lack health services (8), making them important areas for public health interventions and research.

In some CRTs, study endpoints need to be recorded at a site distant from the participant's household—for example,

at a health facility. In this case, researchers face the challenge of correctly determining which of the patients who present to the facility with the condition of interest are cluster residents and to which cluster they belong. Previous CRTs have used systems such as patient-carried cards containing details of each person's cluster residency status, which are presented upon attendance at the facility (9), or regular active home-based follow-up (10). However, both of these approaches can be expensive, time-consuming, and subject to a high failure rate, especially in slums.

Here we describe the development and validation of a novel map book system for ascertaining cluster residency status for persons presenting at clinics in Blantyre, Malawi. All clusters in the parent CRT were located in densely populated slums and had no municipal address system. The objective was to design and validate a method to rapidly, accurately, and reliably categorize a clinic attendee's cluster of residence.

## MATERIALS AND METHODS

### Study design, population, and cluster demarcation

The Hit-TB Hard Study (Hit-TB) (11) and the Convenience in Delivery of ART: Yambirani Pakhomo Study (CONDA-YAPA) (12) are CRTs investigating intensive human immunodeficiency virus and tuberculosis prevention and linkage to antiretroviral therapy (ART) by using community-based testing and treatment strategies in 3 urban slums in Blantyre, Malawi (trial registration numbers: International Standard Randomised Controlled Trial Number Register, ISRCTN02004005; <http://clinicaltrials.gov>, NCT01414413). Two primary care clinics (Ndirande and Chilomoni Health Centers) and a referral facility, which also offers outpatient ART treatment (Queen Elizabeth Central Hospital), serve the study population.

In the parent CRTs, a sample size of 28 clusters was required, and cluster boundaries were based on existing community health worker catchment areas. Circumferential walks of community health worker catchment areas with Global Positioning System (GPS) readings (eTrex Legend HCx; Garmin International, Inc., Olathe, Kansas) demarcated preliminary cluster boundaries. Between April and July 2012, research assistants visited each household within the cluster and enumerated household members. Cluster boundaries were revised after the completion of enumeration to ensure that each cluster had approximately 1,200 adult residents; to ensure that, as far as possible, cluster boundaries followed natural delineators such as roads and rivers; and to minimize contamination (13). The final cluster boundaries were imported into Google Earth Pro (Google, Inc., Mountain View, California) and overlaid onto satellite maps of the study area (GeoEye-1, Eurimage, e-GEOS, Rome, Italy).

In the CONDA-YAPA Study, the primary endpoint is comparison between the study arms of the total proportion of cluster residents who initiate ART during the study period. To estimate the number of ART initiations among cluster residents, the cluster residency status of each ART initiator has to be determined when he or she presents at one of the 3 health facilities.

### Method 1: cluster categorization through description of physical location of residence

We initially investigated whether cluster residency status could be ascertained reliably from a written description of household location. This written description is routinely collected by facility staff from all patients who start ART in Malawi and is recorded in ART registers. Two research assistants extracted household location descriptions from clinic registers for all patients who initiated ART during the month of August 2011 and independently categorized each ART initiator as "cluster resident" (yes/no and which cluster), "out-of-cluster resident," or "uncertain."

### Method 2: cluster categorization with the map book system

For the second method, a map book system was developed. The front cover of the map books featured a high-resolution satellite image of the overall study area and cluster boundaries, annotated with local landmarks (Web Figure 1, available at <http://aje.oxfordjournals.org/>). Subsequent pages showed one cluster boundary on each page, scaled to allow identification of each dwelling (Web Figure 2). Important local landmarks such as health centers, churches, chiefs' residences, social meeting places, and natural features were also indicated on each page.

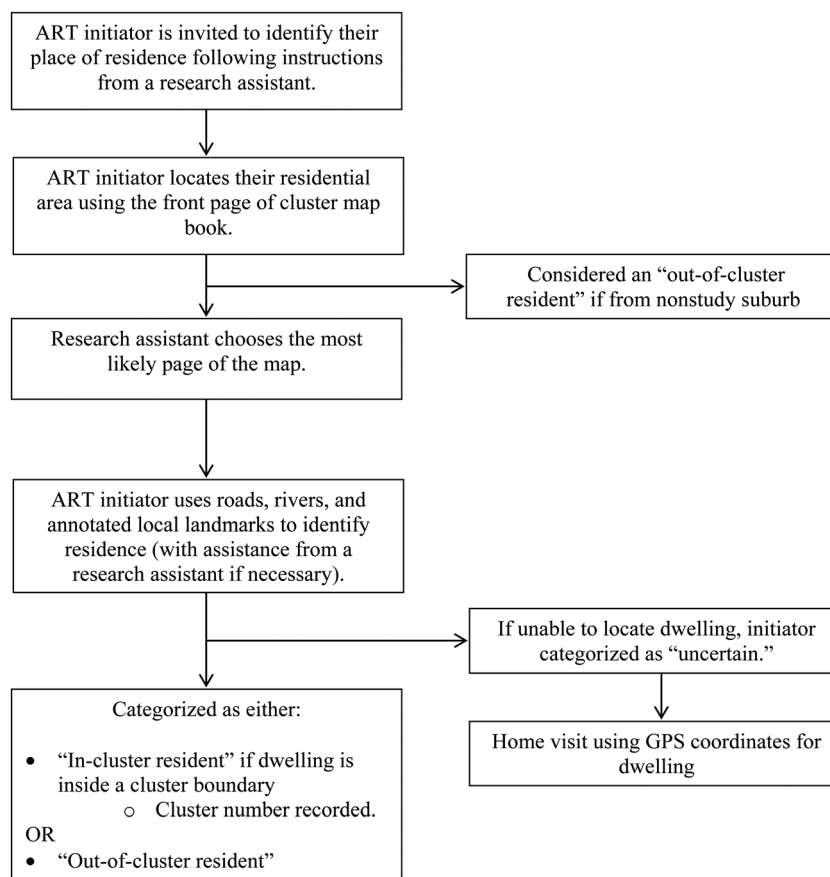
During the month of December 2011, a research assistant and consecutive ART initiators together located the initiator's place of residence (Figure 1). An ART initiator's residence was defined as the place where he or she usually ate meals and slept. Research assistants recorded with a stopwatch the time taken to categorize the cluster residency status of each ART initiator.

### Validation of map book system

A random sample of 10% of ART initiators and a further sample of 10% of ART initiators classified as in-cluster residents were drawn from the study database. ART initiators with uncertain cluster residency status (who were accompanied home to obtain GPS coordinates) were excluded. A research assistant blinded to map book cluster categorization then took a set of GPS coordinates from each ART initiator's household during a home visit.

### Statistical methods

Interrater agreement between the 2 research assistants for Method 1 and agreement between cluster categorization with the map book system and GPS cluster categorization were assessed by using the  $\kappa$  statistic. Bias-corrected 95% confidence intervals for the  $\kappa$  statistic were calculated by using bootstrapped estimates with 1,000 repetitions (14). We estimated that a sample size of 40 was required to give 80% power for the lower limit of the 95% confidence interval for reliability to exceed 0.90, assuming a point estimate of reliability of 0.95 (15). Random sampling and statistical analysis were done in Stata, version 11.2, software (StataCorp LP, College Station, Texas).



**Figure 1.** Map book system for cluster residency status categorization, Blantyre, Malawi, December 2011. ART, antiretroviral therapy; GPS, Global Positioning System.

## Ethics considerations

The research ethics committees of the University of Malawi College of Medicine, the Liverpool School of Tropical Medicine, and the London School of Hygiene and Tropical Medicine granted ethics approval for the parent CRTs.

## RESULTS

### Cluster characteristics

Twenty-eight study clusters were defined through the use of GPS satellite mapping (Web Figure 1). The mean adult cluster population was 1,324 (standard deviation, 107.9), with a mean population density of 0.024 people/m<sup>2</sup> (standard deviation, 0.012). In a pilot study census, 9% of adults were unable to read a letter or newspaper (16).

### Method 1: cluster categorization through description of physical location of residence

In August 2011, the physical locations of residence of 129 adult ART initiators were extracted from the ART initiation registers at the 3 facilities (100% capture). Research assistants

were unable to categorize the cluster residency status of a high proportion of ART initiators: The first research assistant was unable to categorize 41 (31.8%) of 129 ART initiators; the second research assistant was unable to categorize 77 (59.7%) of 129. Overall interrater agreement between the 2 research assistants was 61.5%, with a  $\kappa$  statistic of 0.51 (bias-corrected 95% confidence interval (CI): 0.41, 0.62). After removal of ART initiators who could not be categorized, the agreement and  $\kappa$  statistic were 88.6% and 0.87 (bias-corrected 95% CI: 0.77, 0.97), respectively.

### Method 2: cluster categorization with the map book system

Between December 20, 2011, and January 27, 2012, there were 202 ART initiations at the 3 facilities. Using the map book system in the ART clinics, research assistants assigned 48 ART initiators (23.8%) to a specific study cluster, 147 (72.8%) as outside of all study clusters, and 7 (3.4%) to uncertain cluster status (for GPS determination of final cluster status). The median time to categorize ART initiators' cluster residency status with the map book system was 27 seconds (interquartile range, 20–34 seconds).

The 7 ART initiations with uncertain cluster status on clinic allocation were removed from the database, which left 195 who were included in the sampling frame for GPS validation. After random sampling, 20 ART initiators were selected. There was complete agreement (100%) between map book cluster categorization and GPS cluster categorization, with a  $\kappa$  statistic of 1.00 (bias-corrected 95% CI: 1.00, 1.00).

A further 20 ART initiators categorized as in-cluster residents by the map book system were randomly selected from the study database (after exclusion of persons who had been selected previously). One in-cluster resident was categorized incorrectly as an out-of-cluster resident by the map book system because of close proximity. This gave an agreement and  $\kappa$  statistic of 95.0% and 0.96 (bias-corrected 95% CI: 0.84, 1.00), respectively. The overall agreement between map book categorization and GPS categorization was 97.5%, with a  $\kappa$  statistic of 0.97 (bias-corrected 95% CI: 0.90, 1.00).

## DISCUSSION

This study has shown that cluster residency status can be ascertained rapidly and reliably at facilities distant from community members' place of residence with the use of a simple map book system. The clusters in the study were located in densely populated urban slums with no formal address system and high levels of illiteracy (16). The map book method, if validated in other settings, could be a useful tool for researchers undertaking community-based CRTs in which endpoints are measured at a site distant from participants' places of residence.

This contrasts with our initial finding that categorizing cluster residency status by using physical descriptions of address extracted from ART clinic registers was extremely unreliable, with a high proportion of cases in which research assistants were unable to categorize or agree on cluster residency status. We attribute the poor performance of the physical description method to the lack of systematic addressing, which led clinic staff to record descriptions of household locations in a way that required detailed local knowledge—for instance: “Chilomoni Market, behind the bottle store, ask for Mr. Ngwira.”

Previous community-based studies with the same need to identify cluster residents while at a site distant from the household have used several other methods, including patient-carried referral cards and key-informant systems (9, 17). However, these can be resource intensive and subject to failure. In the present study, the large number of people initiating ART in each clinic and the need for confidentiality made both of these approaches unfeasible. In the validation study reported here, cluster status could be ascertained rapidly in all but a small proportion (3%) of ART initiators, which made home visits necessary only for a few uncertain cases.

GPS technology and up-to-date high-resolution satellite maps were central to the successful development and implementation of this system. In recent years, the increasing availability of free satellite maps of remote and rural areas of low-income countries, as well as software programs for analysis, has resulted in their innovative use in several

research fields (18–20). We took care to include locally important landmarks and to recruit research assistants from within the study wards who were knowledgeable of the study locality and rapidly became familiar with study cluster boundaries.

There were some potential limitations. GPS, maintained by the US government, provides horizontal accuracy within 2.2 meters in more than 95% of measurements (21), which could still lead to occasional misclassification in high-density slums. Of note, on the one instance in which map book categorization differed from GPS categorization, the difference was only 28 meters. Most participants were able to rapidly identify their household in map books annotated with local landmarks. Nevertheless, identification could be more difficult in areas with more transitory populations, where rapid urbanization is occurring, or where there is a predominance of multistory dwellings or forested areas. The varied geography of other settings, such as rural communities, could mean that investigators implementing the map book system might have to modify the scale and layout of cluster maps. Further validation in other settings and with larger sample sizes will be required.

In summary, patients and local research assistants using a low-cost map book system were able to reliably and accurately categorize participants' cluster residency status within the densely populated urban slums of Blantyre. We envisage the map book system, if validated in other settings, as a useful tool for investigators working in several different fields of community-based research.

## ACKNOWLEDGMENTS

Author affiliations: Human Immunodeficiency Virus and Tuberculosis (HIV-TB) Group, Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi (Peter MacPherson, Augustine T. Choko, Deus Thindwa, Rodrick Sambakunsi, Treza Chunda, Kondwani Chavula); Clinical Group, Liverpool School of Tropical Medicine, Liverpool, United Kingdom (Peter MacPherson, S. Bertel Squire, David G. Laloo); Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom (Emily L. Webb); Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, United Kingdom (Elizabeth L. Corbett); Department of Medicine, College of Medicine, University of Malawi, Blantyre, Malawi (Joep J. van Oosterhout); and HIV Department, Ministry of Health of Malawi, Lilongwe, Malawi (Simon D. Makombe).

This work was supported by the Wellcome Trust (grant WT089673).

The authors thank the Antiretroviral Therapy Clinic staff at Ndirande and Chilomoni Health Centers and Queen Elizabeth Central Hospital, Blantyre, Malawi. The authors are also grateful for the support of the Blantyre District Health Office and the HIV Unit of the Ministry of Health of Malawi.

Conflict of interest: none declared.

## REFERENCES

1. Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomization, 1979–1989. *Int J Epidemiol.* 1990;19(4): 795–800.
2. Zelen M. A new design for randomized clinical trials. *N Engl J Med.* 1979;300(22):1242–1245.
3. Bland JM. Cluster randomised trials in the medical literature: two bibliometric surveys. *BMC Med Res Methodol.* 2004;4:21.
4. Hayes RJ, Alexander ND, Bennett S, et al. Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. *Stat Methods Med Res.* 2000;9(2):95–116.
5. Sorensen G, Emmons K, Hunt MK, et al. Implications of the results of community intervention trials. *Annu Rev Public Health.* 1998;19(1):379–416.
6. Torgerson DJ. Contamination in trials: Is cluster randomisation the answer? (review). *BMJ.* 2001;322(7282):355–357.
7. United Nations Human Settlements Programme. *State of the World's Cities 2010/2011: Bridging the Urban Divide.* Nairobi, Kenya: United Nations Human Settlements Programme; 2010.
8. Shetty P. Health care for urban poor falls through the gap (commentary). *Lancet.* 2011;377(9766):627–628.
9. Shekalaghe SA, Drakeley C, van den Bosch S, et al. A cluster-randomized trial of mass drug administration with a gametocytocidal drug combination to interrupt malaria transmission in a low endemic area in Tanzania. *Malar J.* 2011;10:247.
10. Lewycka S, Mwansambo C, Kazembe P, et al. A cluster randomised controlled trial of the community effectiveness of two interventions in rural Malawi to improve health care and to reduce maternal, newborn and infant mortality. *Trials.* 2010;11:88.
11. Corbett EL. Intensified HIV/TB prevention linking home-based HIV testing, including the option of self-testing, with HIV care. London, United Kingdom: Current Controlled Trials, Ltd; 2012. (<http://www.controlled-trials.com/ISRCTN02004005/>). (Accessed August 30, 2012).
12. MacPherson P. Home assessment and initiation of antiretroviral therapy for HIV in Malawi (CONDA-YAPA). Bethesda, MD: ClinicalTrials.gov, US National Institutes of Health; 2012. (<http://clinicaltrials.gov/ct2/show/NCT01414413>). (Accessed August 30, 2012).
13. Hayes RJ, Moulton LH. *Cluster Randomised Trials.* Boca Raton, FL: CRC Press; 2009.
14. Reichenheim ME. Confidence intervals for the kappa statistic. *Stata J.* 2004;4(4):421–428.
15. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med.* 1998;17(1): 101–110.
16. Choko AT, Desmond N, Webb EL, et al. The uptake and accuracy of oral kits for HIV self-testing in high HIV prevalence setting: a cross-sectional feasibility study in Blantyre, Malawi. *PLoS Med.* 2011;8(10):e1001102.
17. Floyd S, Molesworth A, Dube A, et al. Population-level reduction in adult mortality after extension of free anti-retroviral therapy provision into rural areas in northern Malawi. *PLoS ONE.* 2010;5(10):e13499.
18. Stresman GH, Kamanga A, Moono P, et al. A method of active case detection to target reservoirs of asymptomatic malaria and gametocyte carriers in a rural area in Southern Province, Zambia. *Malar J.* 2010;9:265.
19. Vanden Eng JL, Wolkon A, Frolov AS, et al. Use of handheld computers with global positioning systems for probability sampling and data entry in household surveys. *Am J Trop Med Hyg.* 2007;77(2):393–399.
20. Were MC, Kariuki J, Chepng'eno V, et al. Leapfrogging paper-based records using handheld technology: experience from western Kenya. *Stud Health Technol Inform.* 2010; 160(Pt 1):525–529.
21. US Government Federal Aviation Authority. Horizontal position error histogram: 1st January–31st March 2011. Washington, DC: National Coordination Office for Space-Based Positioning, Navigation, and Timing; 2011. (<http://www.gps.gov/systems/gps/performance/accuracy/histogram.png>). (Accessed August 30, 2012).