Observational Research on NCDs in HIV-Positive Populations: Conceptual and Methodological Considerations

Maya Petersen, MD, PhD, Constantin T. Yiannoutsos, PhD,† Amy Justice, MD, PhD,‡ and Matthias Egger, MSc, MD§‖*

Abstract: Noncommunicable diseases (NCDs) account for a growing burden of morbidity and mortality among people living with HIV in low- and middle-income countries (LMICs). HIV infection and antiretroviral therapy interact with NCD risk factors in complex ways, and research into this "web of causation" has so far been largely based on data from high-income countries. However, improving the understanding, treatment, and prevention of NCDs in LMICs requires region-specific evidence. Priority research areas include: (1) defining the burden of NCDs among people living with HIV, (2) understanding the impact of modifiable risk factors, (3) evaluating effective and efficient care strategies at individual and health systems levels, and (4) evaluating cost-effective prevention strategies. Meeting these needs will require observational data, both to inform the design of randomized trials and to replace trials that would be unethical or infeasible. Focusing on Sub-Saharan Africa, we discuss data resources currently available to inform this effort and consider key limitations and methodological challenges. Existing data resources often lack population-based samples; HIV-negative, HIV-positive, and antiretroviral therapy–naive comparison groups; and measurements of key NCD risk factors and outcomes. Other challenges include loss to follow-up, competing risk of death, incomplete outcome ascertainment and measurement of factors affecting clinical decision making, and the need to control for (time-dependent) confounding. We review these challenges and discuss strategies for overcoming them through augmented data collection and appropriate analysis. We conclude with recommendations to improve the quality of data and analyses available to inform the response to HIV and NCD comorbidity in LMICs.

Key Words: HIV/AIDS, Sub-Saharan Africa, resource limited settings, casual inference, loss to follow-up, cohorts

(*J Acquir Immune Defic Syndr* 2014;67:S8–S16)

INTRODUCTION

HIV disease has one necessary cause: infection by HIV. In contrast, many chronic noncommunicable diseases (NCDs), including heart disease, cancer, and chronic pulmonary disease result from the interaction of multiple risk factors, such as smoking, nutrition, physical activity, and genetics.[1] In high-income countries (HICs), many behavioral risk factors occur more commonly in HIV-positive than HIV-negative populations. For example, the prevalence of smoking is considerably higher in people with HIV than in the general population.[2,3] Further, both HIV virus itself and antiretroviral (ARV) drugs used to treat the infection may interact with traditional risk factors to cause NCDs.[4] For example, HIV-induced immune activation leads to inflammation and macrophage recruitment, which in turn is associated with aging, atherosclerosis and cardiovascular disease, and cancer.[5–11] Some ARVs increase levels of low-density lipoprotein cholesterol or reduce insulin sensitivity of peripheral tissues—risk factors also influenced by diet and physical activity.[4,12] Adherence to antiretroviral therapy (ART) affects virologic response and thus immune activation and is in turn associated with lifestyle and psychosocial NCD risk factors.[13,14] Figure 1 illustrates how these complex interactions contribute to cardiovascular disease. Similar causal relationships occur for many other NCDs, including diabetes, liver disease, chronic kidney disease, depression, and a wide range of cancers.[6,10,11,15–18]

Research to date supporting this complex "web of causation" has largely been based on data from HICs. Many causal relationships documented in these countries, such as the inflammatory effects of HIV or potential toxicity of certain ARVs, likely also exist in low- and middle-income

From the *Division of Biostatistics, Berkeley School of Public Health, Berkeley, CA; †Department of Biostatistics, Indiana University R. M. Fairbanks School of Public Health, Indianapolis, IN; ‡Section of General Internal Medicine, Yale University Schools of Medicine and Public Health and Veterans Affairs (VA) Connecticut Healthcare System, West Haven, CT; §Centre for Infectious Disease Epidemiology and Research, School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa; and ‖HIV/AIDS Research Group Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland.

M.P. is a recipient of a Doris Duke Clinical Scientist Development Award. C.Y. is supported by the National Institutes of Health Grant 5U01AI069911-07 (National Institute of Allergy and Infectious Diseases, National Institute of Child Health and Human Development, and National Cancer Institute). A.J. is supported by the National Institutes of Health grants U24-AA020794, U01-AA020790, U01-AA020795, U01-AA020799 (National Institute on Alcohol Abuse and Alcoholism) and in kind by the US Department of Veterans Affairs. M.E. is supported by the National Institutes of Health Grant 5U01AI069924-08 (National Institute of Allergy and Infectious Diseases, National Institute of Child Health and Human Development, and National Cancer Institute), the Swiss National Science Foundation (grants 33CS30_148415, 320030_138490, 33CS30_148522, PDAMP3_137192, 406740_139333, and PDFMP3_137106), and Cancer Research Switzerland.

The authors have no conflicts of interest to disclose.

Correspondence to: Matthias Egger, MSc, MD, HIV/AIDS Research Group, Institute of Social and Preventive Medicine (ISPM), Finkenhubelweg 11, CH-3012 Bern, Switzerland (e-mail: egger@ispm.unibe.ch).

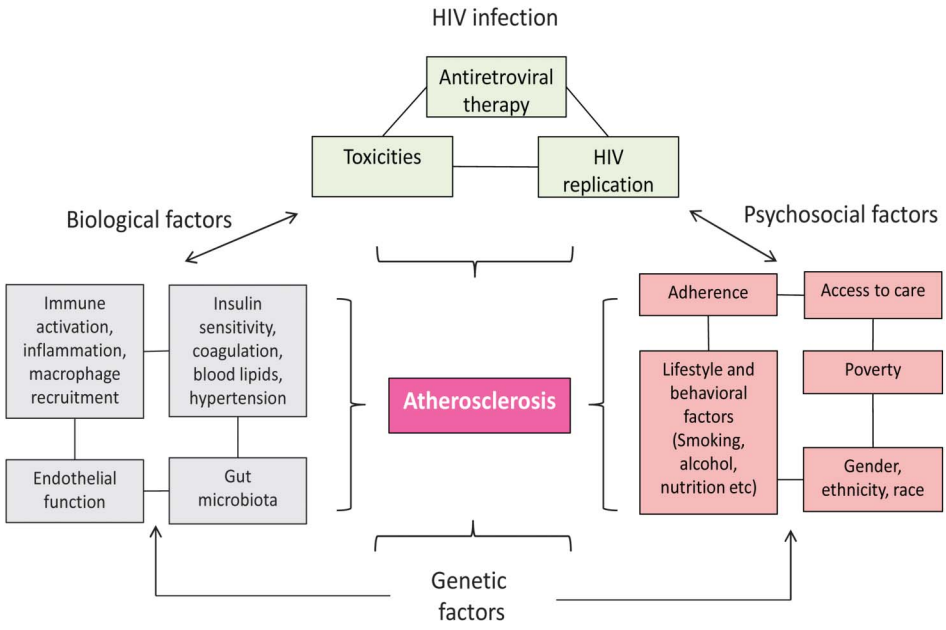

FIGURE 1. Interactions among HIV infection, ART, and different risk factors in cardiovascular disease risk.

countries (LMICs). However, the magnitude of these effects may differ in LMICs due to differing distributions of comorbidities, coinfections, HIV subtypes, and NCD risk factors. LMICs may also differ in their use of more toxic ARVs, and the extent to which NCD risk factors differ between HIV-positive and HIV-negative individuals.[12,19] Improving the understanding, treatment, and prevention of HIV and NCDs in LMICs thus requires region-specific evidence.

Available evidence is insufficient to tackle the challenge of NCDs in the HIV-uninfected population in LMICs and is even more limited for the large number of people living with HIV (PLHIV) in these countries. Priority research areas in the HIV-positive population include: (1) defining the burden of NCDs today and in decades to come, (2) understanding the prevalence and importance of modifiable risk factors, (3) evaluating effective and efficient treatment and care strategies at the individual and health systems levels, and (4) evaluating cost-effective preventive interventions.[19]

Meeting these needs poses significant methodological challenges. Randomized trials will at most allow us to assess the efficacy of a limited number of interventions, and while innovative designs can expand the number of questions amenable to study using randomization,[20,21] many exposures, such as HIV infection itself and access to ART cannot ethically be randomized. Improving prevention and treatment of NCDs in LMICs thus requires us to make the best possible use of observational data. However, because of the complex web of causation of NCDs, disentangling causal effects from confounding or mediating associations in such data is difficult.[22] In this article, we review data resources currently available to inform this effort, focusing on longitudinal observational data from Sub-Saharan Africa (SSA). We then consider key methodological challenges posed by these data and how these challenges can be overcome through augmented data collection and appropriate analysis.

SEARCH STRATEGY

We searched the PubMed database with no language or date specified to identify sources of and gaps in currently available data on NCDs and HIV in SSA and other lower-income settings. We combined MESH and free-text terms describing HIV infection, NCDs, and lower-income settings. For example, we combined [*HIV infections* (Mesh) or *HIV* or *AIDS* or *immunodeficiency virus*] with [*neoplasms* (Mesh) or *neoplasms* or *cancer*] and [*Africa* (Mesh) or *Africa* or *Africa** or *developing countries* (Mesh)]. We used our collections of methodological articles and performed additional searches to identify relevant articles on the methodological challenges discussed in the article and regularly referred to the reviews in the current special issue.

EXISTING DATA RESOURCES AND THEIR LIMITATIONS

The massive scale-up of ART and the evaluation and implementation research that has informed it means that large cohort studies and clinical databases of PLHIV on ART are now available. For example, the International epidemiological Databases to Evaluate AIDS (IeDEA) has assembled data from almost 1 million patients receiving ART across Africa to address clinical and operational research questions around the delivery of ART and clinical outcomes in children, adolescents, and adults.[23–25] Similar IeDEA consortia are providing clinical data from the Americas and Asia.[26,27] The IeDEA data and those collected by other organizations, such as the International Center for AIDS Care and Treatment Programs (ICAP)[28,29] or Médecins Sans Frontières (MSF)[30,31] reflect routine care across a wide range of care settings, including large and small urban and rural clinics run by national health systems or nongovernmental organizations.

There is, however, a dearth of data on NCDs and their risk factors in the HIV-positive population in SSA. In clinical cohort consortia such as IeDEA, data on NCD risk factors and outcomes are not routinely collected, and there are few ART-naive patients with extended follow-up. Further, these data sources include no comparable HIV-negative groups. A recent systematic review of the African Partnership for Chronic Disease Research (APCDR) on cardiovascular risk factors in populations in SSA identified 52 studies, but not a single cohort study that included ART-naive HIV-positive patients and HIV-negative persons.[32] Further, most studies were cross-sectional and based on clinic populations, posing limitations we discuss further below.

Although over 200 Demographic and Health Surveys (DHSs) have been done in general populations since 1985, including over 100 in SSA, the focus of these surveys has generally been on fertility, contraception, and maternal and child health, with no data collected on NCDs.[33] One exception is the Agincourt Health and Demographic Surveillance System in rural northeast South Africa, which has collected data on both HIV status and NCD risk factors for several years.[34] Other DHSs are planning to include NCD risk factors in the future.[35,36] Surveys using the WHO STEP-wise approach to Surveillance (STEPS) for chronic disease risk factors do not generally test for HIV infection.[37] This was the case, for example, in a recent nationwide STEPS survey in Malawi.[38] One notable exception are the STEPS surveys done in rural KwaZulu-Natal, South Africa, within the framework of longitudinal population-based HIV and health surveillance conducted by the Africa Centre for Health and Population Studies.[39,40] Table 1 details typical characteristics of some of the available data and discusses their limitations.

Existing data sources from LMICs therefore pose significant challenges for the study of NCDs and HIV comorbidity, including a lack of population-based samples; absence of HIV-negative, HIV-positive, and ART-naive comparison groups; and incomplete measurement of key NCD risk factors and outcomes. Further, studies of NCDs in PLHIV confront a set of challenges shared by studies of PLHIV in LMICs more generally, including loss to follow-up, incomplete outcome ascertainment, incomplete measurement of key factors affecting clinical decision making, and the need to control for (time-dependent) confounding and validate prognostic models. We review each of these challenges below.

CHALLENGES AND OPPORTUNITIES FOR HIV AND NCD RESEARCH

Clinic Versus Population-based Samples

Although useful for some questions, clinic-based cohorts have significant limitations when the goal is to characterize the underlying population. Differences between those in care and the underlying population may be particularly pronounced in observational studies of NCDs in LMICs for several reasons. First, the socio-demographic characteristics of PLHIV in care differ in important ways from those who are not in care. For example, men, those of lower income and those in more rural areas may be underrepresented in clinical settings.[44] Because risk of NCDs also differs by sex, income, and rural versus urban lifestyles,[45,46] use of a clinic-based sample may cause studies to underestimate or overestimate the disease burden posed by NCDs at the population level and to underestimate or overestimate the importance of specific NCD risk factors. Individuals with more recent infection and better treatment are further overrepresented among cohorts based on patients alive and in care.[47]

Second, many clinics only provide care to PLHIV and are thus unable to identify appropriate HIV-negative comparator samples. Disentangling the effects of HIV infection on NCD incidence from associations due to confounding requires comparisons with demographically and behaviorally similar uninfected individuals. For example, in HICs, smoking, alcohol, and HCV and HBV coinfection are more common among PLHIV. Unless appropriately accounted for, these factors can result in overestimates of the association between HIV infection and a wide range of cancers, liver disease, and neurocognitive disorders.[10,15,18] In LMICs, both magnitudes of such confounding and the key confounding variables may differ because of differences in HIV risk factors. Ideally, HIV-negative comparators would be drawn from the same population as their HIV-positive counterparts.

Third, clinic-based cohorts generally provide limited longitudinal follow-up of ART-naive patients, particularly those with high CD4 counts. Further, patients entering routine clinical care early in the course of HIV disease are likely to differ in important ways from those entering care later. Population-based samples using community-wide HIV testing are thus essential for providing adequate ART-naive comparison groups.

For some research questions, clinic-based cohorts provide an excellent data source. In particular, clinic-based samples are appropriate for analyses that aim to understand and improve clinical outcomes among patients in care. Clinical data systems can be leveraged to measure rich longitudinal data, including on diagnostic tests, routine laboratory monitoring, medication use, and physical findings. Such data are crucial for analyses aimed at developing and evaluating individual and health service level strategies for treatment and care. Examples include developing and evaluating thresholds for intervening on metabolic abnormalities, strategies for modifying ART regimens to prevent alcohol-associated liver disease, adjuvant therapies for the treatment of HIV-associated nephropathy, and strategies for integrated service delivery on a wide range of NCDs.[15,17,18,48] However, realizing this potential requires that diagnostic tests, laboratory monitoring, and medications for treating both HIV and NCDs be available in clinics, and that their indications and use be reliably captured by data systems.

Loss to Follow-up, Competing Risks, and Incomplete Outcome Ascertainment

In SSA, and in many LMICs in other regions, failure to retain PLHIV in care undermines the individual and public health benefits of ART delivery programs. In Africa, loss to follow-up after enrollment often reaches 20%–40% after 2 years.[49–52] High rates of loss to clinic and incomplete outcome

TABLE 1. Examples of Potential Resources for Research on NCDs in HIV-Positive Populations in LMICs

Type of Study	Example	Limitations
Community-based longitudinal studies	The GPC was set up in 1989 to study HIV in rural Southwestern Uganda. The annual surveys were extended in 2010 to collect data on lifestyle factors (smoking, alcohol, physical activity, and diet), blood pressure, and waist and hip circumference; to measure HbA1c and lipid factors; and to perform genetic tests. The 2010/11 survey round included a total of 15,376 persons (80.2% of all eligible), with about 90% of the study population below 50 years of age and an HIV prevalence of 9% among those aged 15–49 yrs[41]	The number of HIV-positive study participants will depend on the HIV prevalence and availability of ART but will typically be relatively small. Detailed data on ART may not be available. The study population will need to be observed for many years and loss to follow-up may be a problem
Routine clinical databases and cohorts	The IeDEA are a global consortium of ART programs in 48 countries and contain data on currently almost 1 million patients.[23] Patients are included as they initiate HIV care at a participating clinic. Data are collected at baseline and each follow-up visit, including the type of ART initiated, and, where available, CD4 counts and HIV-1 plasma RNA levels[23]	NCD risk factors and outcomes are not generally assessed in routine care cohorts and clinical databases of patients on ART. Factors such as loss to follow-up and incomplete engagement in care are often substantial. Most study participants are HIV-positive and most are on ART
Cross-sectional DHSs	The Malawi 2010 DHS was based on a 2-stage stratified sample of 18,000 ever-married women aged 12–49 yrs.[42] In a subsample of ever-married men aged 15–54 yrs who were also interviewed, a group was selected to participate in blood pressure measurements, hemoglobin and blood glucose testing, and height and weight measurements	The DHS evolved from the world fertility and contraception surveys in the 1970s, and the focus continues to be on maternal and child health.[33] Limitations include the cross-sectional nature of the data, the lack of information on risk factors for NCDs, limited or no HIV testing, and the restriction to younger populations
WHO STEPS surveys	A population-based cross-sectional survey was conducted in the Malawian population aged 25–64 yrs,[38] using the WHO STEP-wise approach to chronic disease risk factor Surveillance (STEPS)[37]; a total of 5451 individuals were randomly selected. Data collection was in 3 steps: a questionnaire on demographic and lifestyle data (step 1, 95% had questionnaire data), measurements of height, weight, blood pressure, waist and hip circumference (step 2, 75% had blood pressure measured), and measurement of total cholesterol and fasting blood glucose (step 3, 50% had cholesterol measured)	Limitations include the cross-sectional nature of the survey and the lack of information on the HIV status of participants
The Sustainable East Africa Research on Community Health (SEARCH) trial (NCT01864603)	SEARCH is a cluster randomized trial to evaluate HIV prevention, health, and economic impacts of expanded HIV testing and ART initiation at all CD4 counts, compared with ART delivered according to in-country guidelines. SEARCH will enroll approximately 320,000 individuals (150,000 adults) from 32 communities in Uganda and Kenya. Blood pressure, alcohol intake, random blood glucose, and HIV antibody will be measured longitudinally in population-based cohorts that include both HIV+ and HIV− residents, and measurements will be linked to clinical records. Height and weight will also be measured among children[43]	Although the population size is large, follow-up time is currently planned for 5 yr only, preventing investigation of longer-term ART effects

DHS, Demographic and Health Surveys; GPC, General Population Cohort; IeDEA, International epidemiological Databases to Evaluate AIDS; ART, antiretroviral therapy.

ascertainment also threaten program monitoring and public health evaluations including NCD research.[53–60]

Patients maintained under observation are generally not representative of the entire patient cohort because risk factors for mortality and NCDs, such as male gender and age, are also associated with the probability of being lost to clinic.[61–64] Most analytic methods to reduce bias because of differential loss rely on the frequently unrealistic assumption that outcomes among patients remaining in care are representative of outcomes among patients with similar characteristics who are lost. This assumption breaks down entirely if loss to follow-up is a direct consequence of the unmeasured outcome, posing a particular challenge in African settings, where death accounts for 12%–87% of patients lost to follow-up.[53,55] Studies in which vital status information on patients lost to follow-up was obtained, for example, through linkage with the national death registry in South Africa,[58] or through tracing a random sample of lost patients in Uganda,[55] have highlighted the potential for incomplete death ascertainment to dramatically bias estimates of risk and predictors of mortality.[54,62,63]

Ascertainment of NCD-specific outcomes is even more challenging than ascertainment of all-cause mortality, often requiring extensive diagnostic testing, which may have to be adapted to the Sub-Saharan context. Renal, cancer, and other NCD registries are rare.[10,17] If unmeasured risk factors for a given NCD, or the NCD itself, affect ascertainment, as with death, partial NCD ascertainment will bias estimates of the burden of NCDs at the population level and estimates of the importance of risk factors for NCDs.

Further, even if death ascertainment is reasonably complete, differential mortality patterns can result in misleading conclusions.[57] If deaths are treated as a competing risk,[65,66] the effect of HIV infection on NCDs may be underestimated because PLHIV experience higher mortality rates than demographically and behaviorally similar individuals without HIV, reducing the opportunity for NCDs to occur. Similarly, beneficial effects of ART on NCDs may be obscured because ART reduces mortality. Although analytic approaches that treat death as a censoring event avoid this issue, these approaches estimate a quantity that may be harder to interpret (an effect in the absence of mortality), and they rely on the frequently unrealistic assumption that persons who die would have had similar NCD outcomes had death been prevented as persons (with similar measured characteristics) who survive.

Measuring Exposure and Predictor Variables

Completeness and quality of measurement are also crucial for exposure and predictor variables, and can be undermined by significant obstacles. First, few NCD risk factors are currently recorded in HIV clinical databases. Augmentation of existing databases with measurements of some risk factors, such as body mass index (BMI), hematocrit, or proteinuria may be feasible with minimal investment in some settings with developed infrastructures.[15,17] In other settings, however, the barriers to incorporating even basic laboratory measurements may be substantial.[24]

Second, measurement of self-reported behavioral risk factors poses significant challenges. For example, potential participants in a Kenyan randomized trial who were HIV-positive reported a high prevalence of hazardous and binge drinking when screened by a separate investigator, whereas a much lower prevalence was reported (particularly by women) after screening was transferred to the clinic nurse.[67] On investigation, it was observed that the nurse was asking women: "You don't drink alcohol, do you?" and asking men: "How much alcohol do you drink?", introducing interviewer bias.[68] Clinic staff and patients may be fearful of violating societal norms, and patients may be concerned that accurately reporting use of alcohol, nonadherence to ART, or risky sexual behaviors might jeopardize their access to treatment. Thus, measurements of these risk factors may need to be independently validated, for example, by phosphatidylethanol (PEth) testing for alcohol intake.[69]

Third, to evaluate alternative care strategies, NCD treatments and the key factors driving their use must be measured. To the extent that HIV and NCD comorbidities are treated through distinct health care systems, these data will generally be missing from the HIV clinical databases. Thus, sites in which HIV and NCD care services are integrated have the potential to provide a research platform.[70,71] Such integration to date has been limited in SSA but may accelerate.[6,43,48,72] Further, in settings such as South Africa, with unique patient identifiers and electronic medical records, linkage of regional pharmacy or hospital data with HIV cohort data may provide powerful supplemental data.

Measuring and Controlling for Confounders

Disentangling the effects of HIV infection and ART on NCDs requires controlling for the fact that PLHIV differ systematically from those without HIV, and persons in care and on treatment differ systematically from those off treatment. Although the use of appropriate comparison groups, when available, can help to mitigate confounding resulting from differences between HIV-negative individuals and HIV-positive persons on ART and not on ART, additional analytic adjustment is generally necessary to control for residual differences in risk factors. Further, individuals who receive a treatment are often different from those who do not with respect to factors that influence the outcome. For example, in a study that aims to understand the long-term impact of ART on development of diabetes, a subject's BMI at baseline may affect both initiation of ART and subsequent diabetes risk. Standard techniques, such as contingency tables and multivariable regression, as well as newer targeted data-adaptive approaches can be used to adjust for such baseline confounders.[73,74]

Control for confounding in the study of NCDs and HIV is complicated further because exposures of interest may change over time. In such cases, standard stratification or regression-based approaches to confounding adjustment often break down.[75] For example, time-varying factors such as BMI may affect subsequent ART use and diabetes risk, necessitating their control as confounders.[15] However, these confounders may themselves be on the causal pathway from ART use to diabetes (or simply be affected by previous ART use), preventing standard regression-based adjustment. Similar concerns arise when attempting to disentangle the effects of ongoing viral replication and ART use on depression from the effects of depression on adherence and success of ART.[16,76] When such time-dependent confounding occurs, alternative analytic methods, such as marginal structural models and corresponding inverse weighted or targeted estimators are needed.[75,77,78]

Analytic adjustment for both baseline and time-dependent confounding requires that the key adjustment variables be measured well. Contextual knowledge regarding the web of causality is essential to identifying important confounders for a given question. This understanding can then guide data collection. In particular, key variables affecting treatment decisions, and important determinants or predictors of NCD outcomes, should be captured.

Prognostic and Mathematical Modeling

A prognostic model returns a predicted risk of some outcome for an individual, based on that patient's values of multiple predictors. Missing data, the modeling of continuous predictor variables, the complexity of the model, and the checking of model assumptions are important issues in this context.[79] A range of methods are available for constructing prognostic models, including multivariable regression models and a range of "machine-learning" approaches that automate the processes of selecting predictor variables, modeling the functional form of relationships between outcomes and continuous predictors, and selecting an optimal degree of model complexity.[80,81]

Formalized risk assessment using a prognostic model (prediction index or rule) can improve the efficiency of clinical care and research.[79] Examples of prognostic models in HIV-positive populations include the model of mortality in patients starting ART in SSA developed by IeDEA[24] (see www.iedea-sa.org for web calculator), and the Veterans Aging Cohort Study (VACS) Index (http://vacs.med.yale.edu).[82,83] Potential applications to the prevention and treatment of NCDs in LMICs include targeted screening for metabolic risk factors and interventions such as directing lipid lowering drugs to individuals at higher risk,[15,84] determining appropriate intervals for clinical follow-up, assessing response to interventions, and planning for resource requirements.

Models do not need to include all possible risk factors or biomarkers. Once several informative variables are included, additional factors may not substantially improve the predictive accuracy of the index. For example, in the prognostic model of patients starting ART in SSA,[24] the inclusion of either the CD4 cell count or the hemoglobin and total lymphocyte count resulted in a model of similar prognostic power that was not improved by including all 3 variables.

It is essential to demonstrate that a prognostic model provides valid predictions outside the dataset and narrow context in which it was developed. Cross-validation is a method that uses multiple splits of the database, allowing performance to be evaluated with data not used for building the prognostic model.[80,85] If the model is used in distinct geographical areas, ideally external validation would also be performed in these regions (geographic external validation).[86] In assessing performance, a range of metrics is available. For example, discrimination and calibration can be assessed by calculating the c-statistic, risk, and net reclassification index, and plotting receiver operating curve and calibration curves.[87,88] Finally, when prognostic models are used as medical interventions, their impact on health outcomes should ideally be evaluated, either in randomized control trials or using observational data.[79]

Mathematical modeling complements and informs randomized control trials and observational research, providing a means of exploring potential impacts of many different intervention strategies under a range of assumptions. Mathematical models have been widely used both in the field of HIV and NCDs, and will likely play a role in answering several questions pertinent to NCDs in PLHIV. The HIV Modeling Consortium provides a forum for identifying questions, sharing models, data and expertise, and rigorous review of modeling research.[89]

BUILDING RESEARCH CAPACITY

Expertise in study design, data collection and management, and statistical analysis is needed to effectively carry out the research discussed in this article.[90] However, such capacity is lacking in many settings in SSA. Although substantial investments have been made in training a labor force with the methodological expertise needed to address the challenge of NCDs,[91,92] expanded training opportunities for local investigators, both in the form of on-the-job training and through master's and PhD level courses in biostatistics, epidemiology, and information sciences, are often lacking in LMICs. Further, development of a mentored and well-supported research career path for LMIC investigators is needed to ensure that investments in graduate-level training translate into sustainable improvements in LMIC research capacity. Both South–South collaborations such as the ALPHA network (Analyzing Longitudinal Population-based HIV/AIDS data on Africa),[93] and collaborative research partnerships between institutions in low-income and HICs can play a role.[94] Ultimately, however, building sustainable research capacity requires a long-term commitment from donors, local governments, and academic institutions in both low-income and HICs.[90,95]

CONCLUSIONS AND RECOMMENDATIONS

We propose several recommendations (Box 1) to improve the quality of data and analyses available to inform the response to HIV and NCD comorbidity in LMICs. First, published guidance for the transparent conduct and reporting of observational research should be considered.[79,96,97] Second, optimal use of existing data sources should be made, for example by linking routine HIV clinical databases to laboratory databases, cancer registries, pharmacy and hospitalization records, and electronic medical record systems where available. In addition, for NCDs with sufficiently high prevalence and incidence, substudies can be added to large prospective studies that are addressing other questions. Third, data collection in existing studies and clinical databases should be strategically expanded to obtain high-quality data on NCDs, key risk factors such as smoking and hypertension, and key variables that influence treatment decisions. Fourth, uninfected comparison groups should be chosen carefully, taking into account established risk factors. Fifth, analyses of observational data should take into account key methodological challenges. Sixth, collaborative research is required to overcome the limited power of single studies and to validate prognostic and mathematical models. Finally, local methodological research capacity should be strengthened.

In conclusion, although this article has focused on methodological challenges to observational research, many of

Box 1. Recommendations Aimed at Investigators Involved in Observational Research on NCD in HIV-Positive Populations

1. Consider relevant guidance on the conduct and reporting of observational research, for example:
   - PROGnosis RESearch Strategy (PROGRESS) recommendations[79,96]
   - Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations.[97]
2. Make optimal use of existing data sources:
   - Facilitate record linkage between HIV clinical databases and disease, pharmacy, hospital, and mortality registries
   - Consider adding NCD components to other large prospective studies such as Treatment as Prevention (TasP) trials.
3. Enhance recruitment and data collection in ongoing studies and clinical databases, in sentinel sites, or all sites:
   - Recruit behaviorally and demographically similar untreated HIV-positive and HIV-negative individuals
   - Assess core risk factors, including HIV infection, hypertension and smoking, and key variables that influence clinical decision making
   - Trace individuals "lost to follow-up" to ascertain outcomes.
4. Justify choice of uninfected comparator groups carefully:
   - Consider health behaviors, demographics, and other important established risk factors.
5. Analyze data considering key methodological challenges, using appropriate statistical techniques to minimize their impact, and acknowledging potential for residual bias:
   - Selection bias, including that resulting from clinic versus population-based samples
   - Confounding, for example, because of differences between HIV-positive and HIV-negative populations
   - Time-dependent confounding due to the impact of time-varying patient characteristics affected by previous treatment on treatment decisions over time
   - Informative missingness of data, including loss to follow-up
   - Measurement error in predictors, exposures, and confounders
   - Competing risks
   - Validation of prognostic and mathematical models.
6. Establish, facilitate, and participate in collaborative research, for example:
   - Meta-analyses of individual participant data to overcome the limited power of single studies
   - Validation of prognostic models in independent data
   - Comparisons of predictions between different mathematical models and real-world data.
7. Contribute to building local research capacity, for example:
   - Provide on-the-job and formal training opportunities for clinician-scientists, epidemiologists, statisticians, and data managers
   - Create master's and PhD opportunities linked to the research
   - Lobby donors and governments to build long-term partnerships and strengthen African research institutions.

these same challenges also directly undermine the delivery of quality clinical care and public health interventions. Thus, investment in overcoming these challenges, including improving engagement in care, integrating health data systems, and measuring key NCD risk factors has the potential to profoundly benefit the health of both HIV-positive and HIV-negative populations in SSA and other LMICs.

ACKNOWLEDGMENTS

The authors are grateful to the editorial team, Paolo Miotti, Nalini Anand, Sten Vermund, Christine Harmston, Lydia Kline, and the two external reviewers, Till Bärnighausen and Elvin Geng, for helpful comments on earlier drafts of this article. The authors also thank Julia Bohlius, Cindy Zahnd, and Kalyani Thyagarajan for help with literature searches.

REFERENCES

1. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. *Nat Rev Genet.* 2006;7:812–820.
2. Glass TR, Ungsedhapand C, Wolbers M, et al. Prevalence of risk factors for cardiovascular disease in HIV-infected patients over time: the Swiss HIV Cohort Study. *HIV Med.* 2006;7:404–410.
3. Tesoriero JM, Gieryic SM, Carrascal A, et al. Smoking among HIV positive New Yorkers: prevalence, frequency, and opportunities for cessation. *AIDS Behav.* 2010;14:824–835.
4. Currier JS, Lundgren JD, Carr A, et al. Epidemiological evidence for cardiovascular disease in HIV-infected patients and relationship to highly active antiretroviral therapy. *Circulation*. 2008;118:e29–35.
5. Hsue PY, Deeks SG, Hunt PW. Immunologic basis of cardiovascular disease in HIV-infected adults. *J Infect Dis*. 2012;205(suppl 3):S375–S382.
6. Deeks SG, Lewin SR, Havlir DV. The end of AIDS: HIV infection as a chronic disease. *Lancet*. 2013;382:1525–1533.
7. Deeks SG, Tracy R, Douek DC. Systemic effects of inflammation on health during chronic HIV infection. *Immunity*. 2013;39:633–645.
8. Borges ÁH, Dubrow R, Silverberg MJ. Factors contributing to risk for cancer among HIV-infected individuals, and evidence that earlier combination antiretroviral therapy will alter this risk. *Curr Opin HIV AIDS.* 2014;9:34–40.
9. High KP, Brennan-Ing M, Clifford DB, et al. HIV and aging: state of knowledge and areas of critical need for research. A report to the NIH Office of AIDS Research by the HIV and Aging Working Group. *J Acquir Immune Defic Syndr.* 2012;60(suppl 1):S1–S18.
10. Adebamowo CA, Casper C, Bhatia K, et al. Challenges in the detection, prevention, and treatment of HIV-associated malignancies in low- and middle-income countries in Africa. *J Acquir Immune Defic Syndr.* 2014;67(suppl 1):S17–S26.
11. Bloomfield GS, Khazanie P, Morris A, et al. HIV and noncommunicable cardiovascular and pulmonary diseases in low- and middle-income countries in the ART era: what we know and best directions for future research. *J Acquir Immune Defic Syndr.* 2014;67(suppl 1):S40–S53.
12. Ebrahim O, Mazanderani AH. Recent developments in HIV treatment and their dissemination in poor countries. *Infect Dis Rep.* 2013;5(suppl 1):e2.

13. Press N, Tyndall MW, Wood E, et al. Virologic and immunologic response, clinical progression, and highly active antiretroviral therapy adherence. *J Acquir Immune Defic Syndr.* 2002;31(suppl 3):S112–S117.
14. Berg CJ, Michelson SE, Safren SA. Behavioral aspects of HIV care: adherence, depression, substance use, and HIV-transmission behaviors. *Infect Dis Clin North Am.* 2007;21:181–200.
15. Ali MK, Magee MJ, Dave JA, et al. HIV and metabolic, body, and bone disorders: what we know from low- and middle-income countries. *J Acquir Immune Defic Syndr.* 2014;67(suppl 1):S27–S39.
16. Chibanda D, Benjamin L, Weiss HA, et al. Mental, neurological, and substance use disorders in people living with HIV/AIDS in low- and middle-income countries. *J Acquir Immune Defic Syndr.* 2014;67(suppl 1): S54–S67.
17. Kalyesubula R, Wearne N, Semitala FC, et al. HIV-associated renal and genitourinary comorbidities in Africa. *J Acquir Immune Defic Syndr.* 2014;67(suppl 1):S68–S78.
18. Kelly P, Saloojee H, Chen JY, et al. Noncommunicable diseases in HIV infection in low- and middle-income countries: gastrointestinal, hepatic, and nutritional aspects. *J Acquir Immune Defic Syndr.* 2014;67(suppl 1): S79–S86.
19. Ebrahim S, Pearce N, Smeeth L, et al. Tackling non-communicable diseases in low- and middle-income countries: is the evidence from high-income countries all we need? *PLoS Med.* 2013;10:e1001377.
20. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Stat Med.* 2005;24:1455–1481.
21. Rosenblum M, Van der Laan MJ. Optimizing randomized trial designs to distinguish which subpopulations benefit from treatment. *Biometrika.* 2011;98:845–860.
22. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health.* 2000;21:121–145.
23. Egger M, Ekouevi DK, Williams C, et al. Cohort profile: the international epidemiological databases to evaluate AIDS (IeDEA) in sub-Saharan Africa. *Int J Epidemiol.* 2012;41:1256–1264.
24. May M, Boulle A, Phiri S, et al. Prognosis of patients with HIV-1 infection starting antiretroviral therapy in sub-Saharan Africa: a collaborative analysis of scale-up programmes. *Lancet.* 2010;376:449–457.
25. Davies MA. Prognosis of children with HIV-1 infection starting antiretroviral therapy in Southern Africa: a collaborative analysis of treatment programs. *Pediatr Infect Dis J.* 2014;33:608–616.
26. McGowan CC, Cahn P, Gotuzzo E, et al. Cohort Profile: Caribbean, Central and south America network for HIV research (CCASAnet) collaboration within the international epidemiologic databases to evaluate AIDS (IeDEA) programme. *Int J Epidemiol.* 2007;36:969–976.
27. Wright S, Boyd MA, Yunihastuti E, et al. Rates and factors associated with major modifications to first-line combination antiretroviral therapy: results from the Asia-Pacific region. *PLoS One.* 2013;8:e64902.
28. Rabkin M, El-Sadr W. *Saving mothers, saving families: the MTCT-plus initiative. Perspectives and Practice in Antiretroviral Treatment*. Geneva, Switzerland: WHO; 2003.
29. Rabkin M, Kruk ME, El-Sadr WM. HIV, aging and continuity care: strengthening health systems to support services for noncommunicable diseases in low-income countries. *AIDS.* 2012;26(suppl 1):S77–S83.
30. Ferradini L, Jeannin A, Pinoges L, et al. Scaling up of highly active antiretroviral therapy in a rural district of Malawi: an effectiveness assessment. *Lancet.* 2006;367:1335–1342.
31. Ford N, Calmy A, Mills EJ. The first decade of antiretroviral therapy in Africa. *Globalization and Health.* 2011;7:33.
32. Dillon D, Gurdasani D, Riha J, et al. Association of HIV and ART with cardiometabolic risk factors in sub-Saharan Africa: a systematic review and meta-analysis. *Int J Epidemiol.* 2013;42:1754–1771.
33. Short Fabic M, Choi Y, Bird S. A systematic review of demographic and health surveys: data availability and utilization for research. *Bull World Health Organ.* 2012;90:604–612.
34. Kahn K, Collinson MA, Gómez-Olivé FX, et al. Profile: Agincourt health and socio-demographic surveillance system. *Int J Epidemiol.* 2012;41: 988–1001.
35. Odhiambo FO, Laserson KF, Sewe M, et al. Profile: the KEMRI/CDC health and demographic surveillance system—Western Kenya. *Int J Epidemiol.* 2012;41:977–987.
36. Crampin AC, Dube A, Mboma S, et al. Profile: the Karonga health and demographic surveillance system. *Int J Epidemiol.* 2012;41:676–685.
37. *Noncommunicable Diseases and Mental Health Cluster (2005) WHO STEPS Surveillance Manual: The WHO STEPwise Approach to Chronic Disease Risk Factor Surveillance*. Geneva, Switzerland: World Health Organization; 2005.
38. Msyamboza KP, Ngwira B, Dzowela T, et al. The burden of selected chronic non-communicable diseases and their risk factors in Malawi: nationwide STEPS survey. *PLoS One.* 2011;6:e20316.
39. Bärnighausen T, Welz T, Hosegood V, et al. Hiding in the shadows of the HIV epidemic: obesity and hypertension in a rural population with very high HIV prevalence in South Africa. *J Hum Hypertens.* 2008;22:236–239.
40. Malaza A, Mossong J, Bärnighausen T, et al. Hypertension and obesity in adults living in a high HIV prevalence rural area in South Africa. *PLoS One.* 2012;7:e47761.
41. Asiki G, Murphy G, Nakiyingi-Miiro J, et al. The general population cohort in rural south-western Uganda: a platform for communicable and non-communicable disease studies. *Int J Epidemiol.* 2013;42:129–141.
42. National Statistical Office (NSO) and ICF Macro. *Malawi. Demographic and Health Survey 2010*. Zomba, Malawi: NSO and ICF Macro; 2011.
43. Chamie G, Kwarisiima D, Clark TD, et al. Leveraging rapid community-based HIV testing campaigns for non-communicable diseases in rural Uganda. *PLoS One.* 2012;7:e43400.
44. Braitstein P, Boulle A, Nash D, et al. Gender and the use of antiretroviral treatment in resource-constrained settings: findings from a multicenter collaboration. *J Womens Health (Larchmt).* 2008;17:47–55.
45. Miranda JJ, Gilman RH, Smeeth L. Differences in cardiovascular risk factors in rural, urban and rural-to-urban migrants in Peru. *Heart.* 2011; 97:787–796.
46. Danaei G, Singh GM, Paciorek CJ, et al. The global cardiovascular risk transition: associations of four metabolic risk factors with national income, urbanization, and Western diet in 1980 and 2008. *Circulation.* 2013;127:1493–1502, 1502e1491-1498.
47. Hoover D, Munoz A, Carey V, et al. The unseen sample in cohort studies: estimation of its size and effect. *Stat Med.* 1991;10:1993–2003.
48. Hyle EP, Naidoo K, Su AE, et al. HIV, tuberculosis and noncommunicable diseases: what is known about the costs, effects, and cost-effectiveness of integrated care? *J Acquir Immune Defic Syndr.* 2014; 67(suppl 1):S87–S95.
49. Fox MP, Rosen S. Patient retention in antiretroviral therapy programs up to three years on treatment in sub-Saharan Africa, 2007-2009: systematic review. *Trop Med Int Health.* 2010;15(suppl 1):1–16.
50. Rosen S, Fox MP. Retention in HIV care between testing and treatment in sub-Saharan Africa: a systematic review. *PLoS Med.* 2011;8: e1001056.
51. Rosen S, Fox MP, Gill CJ. Patient retention in antiretroviral therapy programs in sub-Saharan Africa: a systematic review. *PLoS Med.* 2007;4:e298.
52. Tassie JM, Baijal P, Vitoria MA, et al. Trends in retention on antiretroviral therapy in national programs in low-income and middle-income countries. *J Acquir Immune Defic Syndr.* 2010;54:437–441.
53. Brinkhof MW, Pujades-Rodriguez M, Egger M. Mortality of patients lost to follow-up in antiretroviral treatment programmes in resource-limited settings: systematic review and meta-analysis. *PLoS One.* 2009;4:e5790.
54. Egger M, Spycher BD, Sidle J, et al. Correcting mortality for loss to follow-up: a nomogram applied to antiretroviral treatment programmes in sub-Saharan Africa. *PLoS Med.* 2011;8:e1000390.
55. Geng EH, Emenyonu N, Bwana MB, et al. Sampling-based approach to determining outcomes of patients lost to follow-up in antiretroviral therapy scale-up programs in Africa. *JAMA.* 2008;300:506–507.
56. Geng EH, Glidden DV, Bangsberg DR, et al. A causal framework for understanding the effect of losses to follow-up on epidemiologic analyses in clinic-based cohorts: the case of HIV-infected patients on antiretroviral therapy in Africa. *Am J Epidemiol.* 2012;175:1080–1087.
57. Schoni-Affolter F, Keiser O, Mwango A, et al. Estimating loss to follow-up in HIV-infected patients on antiretroviral therapy: the effect of the competing risk of death in Zambia and Switzerland. *PLoS One.* 2011;6:e27919.
58. Van Cutsem G, Ford N, Hildebrand K, et al. Correcting for mortality among patients lost to follow up on antiretroviral therapy in South Africa: a cohort analysis. *PLoS One.* 2011;6:e14684.
59. Shepherd BE, Blevins M, Vaz LM, et al. Impact of definitions of loss to follow-up on estimates of retention, disease progression, and mortality: application to an HIV program in Mozambique. *Am J Epidemiol.* 2013; 178:819–828.

60. Grimsrud AT, Cornell M, Egger M, et al. Impact of definitions of loss to follow-up (LTFU) in antiretroviral therapy program evaluation: variation in the definition can have an appreciable impact on estimated proportions of LTFU. *J Clin Epidemiol.* 2013;66:1006–1013.
61. Stringer JS, Zulu I, Levy J, et al. Rapid scale-up of antiretroviral therapy at primary care sites in Zambia: feasibility and early outcomes. *JAMA*. 2006;296:782–793.
62. Yiannoutsos CT, An MW, Frangakis CE, et al. Sampling-based approaches to improve estimation of mortality among patient dropouts: experience from a large PEPFAR-funded program in Western Kenya. *PLoS One.* 2008;3:e3843.
63. Towards Universal Access. Scaling up priority HIV/AIDS interventions in the health sector. 2009 progress reportA1-World Health Organization. 2009. Available at: http://www.who.int/hiv/pub/2009progressreport/en/index.html. Accessed June 29, 2014.
64. Schomaker M, Gsponer T, Estill J, et al. Non-ignorable loss to follow-up: correcting mortality estimates based on additional outcome ascertainment. *Stat Med.* 2014;33:129–142.
65. Lau B, Gange SJ, Moore RD. Risk of non-AIDS-related mortality may exceed risk of AIDS-related mortality among individuals enrolling into care with CD4+ counts greater than 200 cells/mm3. *J Acquir Immune Defic Syndr.* 2007;44:179–187.
66. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc.* 1999;94:496–509.
67. Papas RK, Sidle JE, Gakinya BN, et al. Treatment outcomes of a stage 1 cognitive-behavioral trial to reduce alcohol use among human immunodeficiency virus-infected out-patients in western Kenya. *Addiction*. 2011; 106:2156–2166.
68. Wynder EL. Investigator bias and interviewer bias: the problem of reporting systematic error in epidemiology. *J Clin Epidemiol.* 1994;47:825–827.
69. Aradottir S, Asanovska G, Gjerss S, et al. PHosphatidylethanol (PEth) concentrations in blood are correlated to reported alcohol intake in alcohol-dependent patients. *Alcohol Alcohol.* 2006;41:431–437.
70. Fultz SL, Skanderson M, Mole LA, et al. Development and verification of a"virtual" cohort using the National VA Health Information System. *Med Care.* 2006;44:S25–S30.
71. Justice AC, Erdos J, Brandt C, et al. The Veterans Affairs Healthcare System: a unique laboratory for observational and interventional research. *Med Care.* 2006;44:S7–S12.
72. Geneau R, Hallen G. Toward a systemic research agenda for addressing the joint epidemics of HIV/AIDS and noncommunicable diseases. *AIDS.* 2012;26:S7–S10.
73. Bembom O, Petersen ML, Rhee SY, et al. Biomarker discovery using targeted maximum-likelihood estimation: application to the treatment of antiretroviral-resistant HIV infection. *Stat Med.* 2009;28:152–172.
74. van der Laan M, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Berlin Heidelberg, New York: Springer; 2011.
75. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
76. Tsai AC, Weiser SD, Petersen ML, et al. A marginal structural model to estimate the causal effect of antidepressant medication treatment on viral suppression among homeless and marginally housed persons with HIV. *Arch Gen Psychiatry.* 2010;67(12):1282–1290.
77. Petersen ML, Wang Y, van der Laan MJ, et al. Assessing the effectiveness of antiretroviral adherence interventions. Using marginal structural models to replicate the findings of randomized controlled trials. *J Acquir Immune Defic Syndr.* 2006;43(suppl 1):S96–S103.
78. Petersen ML, Schwab J, Gruber S, et al. Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *J Causal Inference.* 2014;[ePub ahead of print, June 18, 2014].
79. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med.* 2013;10:e1001381.
80. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York, NY: Springer-Verlag; 2009.
81. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol.* 2007;6:Article25.
82. Justice AC, Modur SP, Tate JP, et al. Predictive accuracy of the Veterans Aging Cohort Study index for mortality with HIV infection: a North American cross cohort analysis. *J Acquir Immune Defic Syndr*. 2013;62:149–163.
83. Tate JP, Justice AC, Hughes MD, et al. An internationally generalizable risk index for mortality after one year of antiretroviral therapy. *AIDS*. 2013;27:563–572.
84. Greving JP, Visseren FL, de Wit GA, et al. Statin treatment for primary prevention of vascular disease: whom to treat? Cost-effectiveness analysis. *BMJ*. 2011;342:d1672.
85. van der Laan M, Dudoit S, Keles S. Asymptotic optimality of likelihood-based cross-validation. *Stat Appl Genet Mol Biol.* 2004;3:Article4.
86. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 1999;130:515–524.
87. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. *Ann Intern Med.* 2009;150:795–802.
88. Pencina MJ, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* 2008;27:157–172.
89. HIV Modelling Consortium Treatment as Prevention Editorial Writing Group. HIV treatment as prevention: models, data, and questions–towards evidence-based decision-making. *PLoS Med.* 2012;9:e1001259.
90. Lansang MA, Dennis R. Building capacity in health research in the developing world. *Bull World Health Organ.* 2004;82:764–770.
91. Breman JG, Bridbord K, Kupfer LE, et al. Global health: the Fogarty International Center, National Institutes of Health: vision and mission, programs, and accomplishments. *Infect Disease Clinics North America.* 2011;25:511.
92. Heimburger DC, Carothers CL, Gardner P, et al. Nurturing the global workforce in clinical research: the National Institutes of Health Fogarty International Clinical Scholars and Fellows Program. *Am J Trop Med Hyg.* 2011;85:971–8.
93. Maher D, Biraro S, Hosegood V, et al. Translating global health research aims into action: the example of the ALPHA network. *Trop Med Int Health.* 2010;15:321–328.
94. Einterz RM, Kimaiyo S, Mengech HN, et al. Responding to the HIV pandemic: the power of an academic medical partnership. *Acad Med.* 2007;82:812–818.
95. Palen J, El-Sadr W, Phoya A, et al. PEPFAR, health system strengthening, and promoting sustainability and country ownership. *J Acquir Immune Defic Syndr*. 2012;60(suppl 3):S113–S119.
96. Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*. 2013;346:e5595.
97. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med.* 2007;4:e297.