Rethinking methods for patient costs in economic evaluations in LMIC

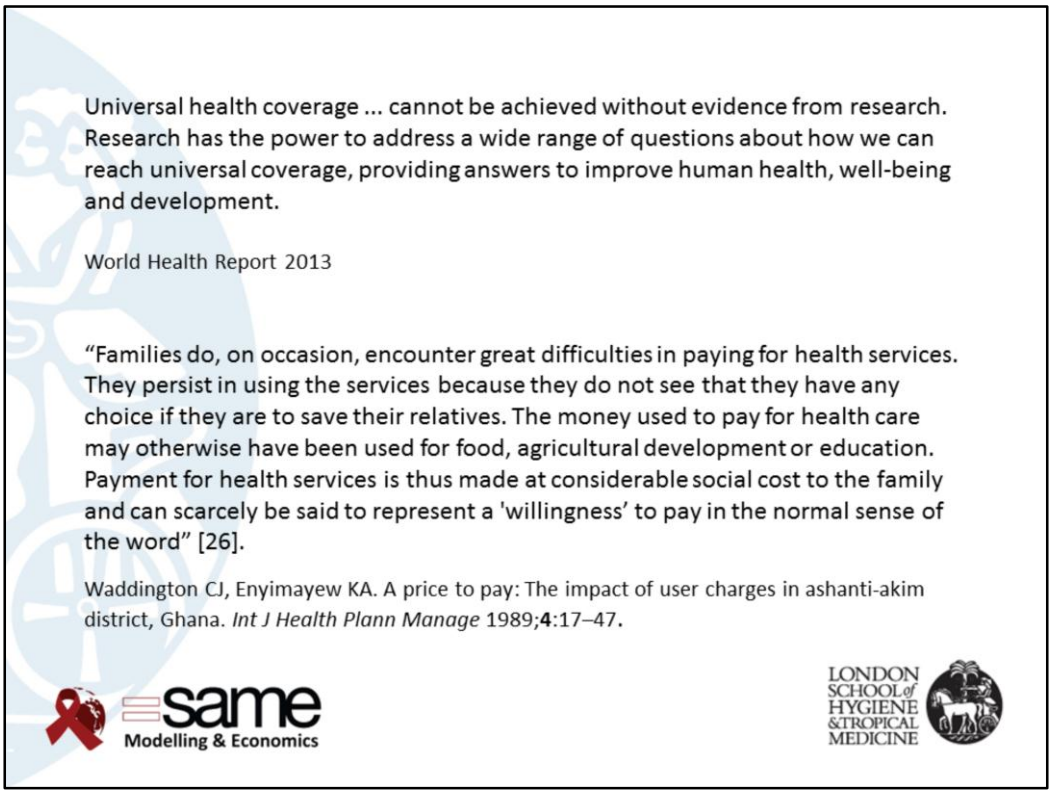
Presented at the 2015 International Health Economics Association Congress, Milan, Italy
Session Title: Methodological Issues of Health Care Cost Estimates in LMIC Settings;
Tuesday, 4:30 p.m.-5:45 p.m.

Sedona Sweeney[1*], Anna Vassall[1], Nicola Foster[3], Victoria Simms[1], Patrick Ilboudo[2], Godfather Kimaro[4], Don Mudzengi[5], Lorna Guinness[1]

[1]London School of Hygiene & Tropical Medicine, United Kingdom,
[2] Centre Muraz, Burkina Faso
[3]University of Cape Town, South Africa
[4] National Institute for Medical Research, Tanzania
[5]Aurum Institute, South Africa

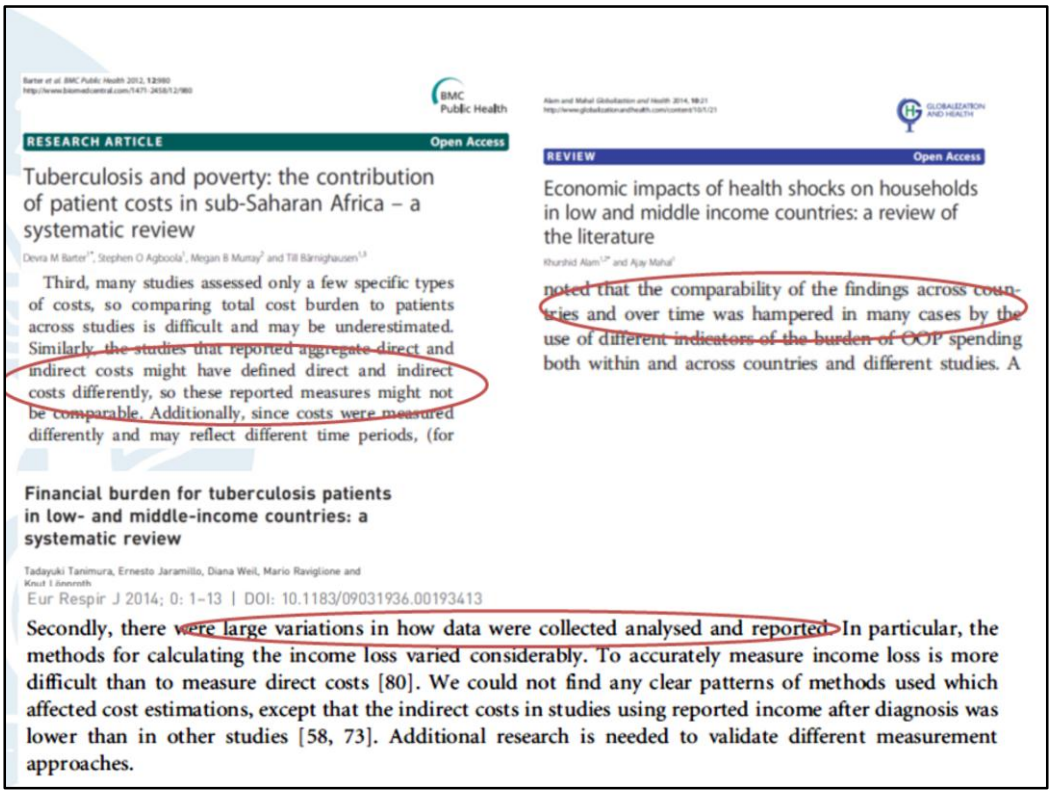
same
Modelling & Economics
http://blogs.lshtm.ac.uk/samegroup/

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE



We are seeing renewed interest in measuring the impact of health spending on poverty and economic vulnerability in the context of the current drive for universal health care. Research on catastrophic or impoverishing spending often uses national-level data, either from national health accounts or from naitonal household surveys such as the Living Standards Measurement Survey or World Health Survey.

However, these national-level household budget surveys can't accurately represent the impact of a specific health intervention on poverty. Research evaluating particular health interventions therefore usually use data from a smaller randomized controlled trial or other similar piece of research.



RESEARCH ARTICLE Open Access

Tuberculosis and poverty: the contribution of patient costs in sub-Saharan Africa – a systematic review

Devra M Barter[1*], Stephen O Agboola[1], Megan B Murray[2] and Till Bärnighausen[1,3]

Third, many studies assessed only a few specific types of costs, so comparing total cost burden to patients across studies is difficult and may be underestimated. Similarly, the studies that reported aggregate direct and indirect costs might have defined direct and indirect costs differently, so these reported measures might not be comparable. Additionally, since costs were measured differently and may reflect different time periods, (for

REVIEW Open Access

Economic impacts of health shocks on households in low and middle income countries: a review of the literature

Khurshid Alam[1,2*] and Ajay Mahal[1]

noted that the comparability of the findings across countries and over time was hampered in many cases by the use of different indicators of the burden of OOP spending both within and across countries and different studies. A

Financial burden for tuberculosis patients in low- and middle-income countries: a systematic review

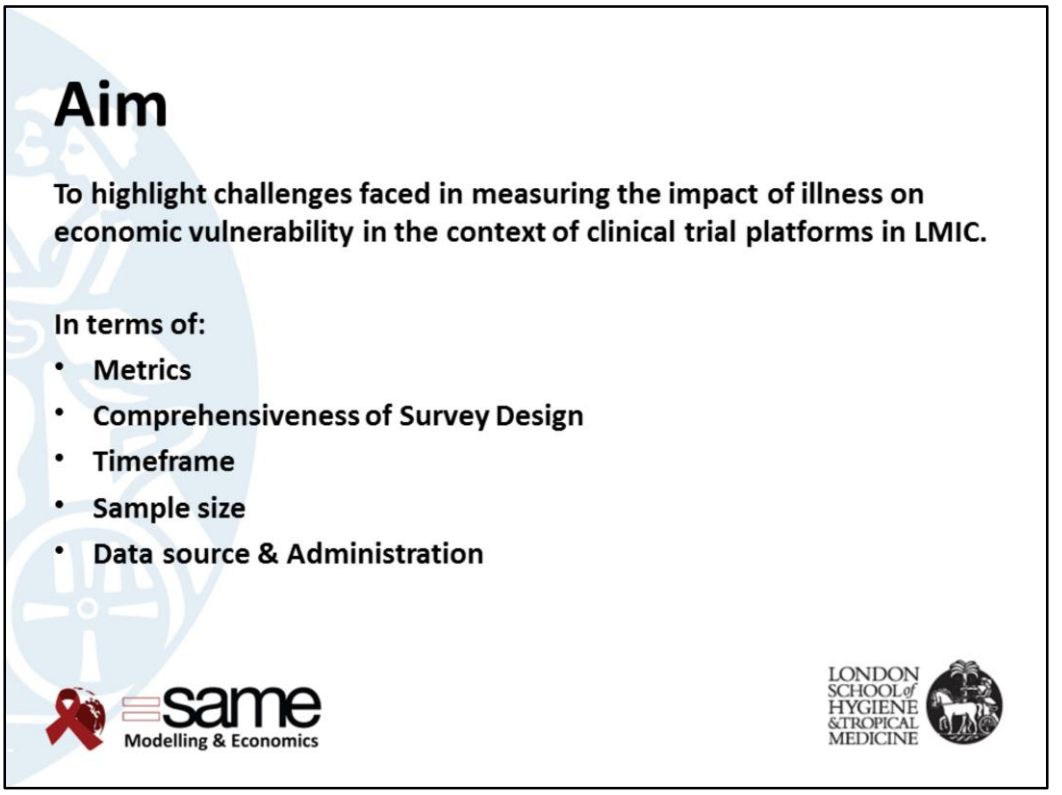
Tadayuki Tanimura, Ernesto Jaramillo, Diana Weil, Mario Raviglione and Knut Lönnroth

Eur Respir J 2014; 0: 1–13 | DOI: 10.1183/09031936.00193413

Secondly, there were large variations in how data were collected analysed and reported. In particular, the methods for calculating the income loss varied considerably. To accurately measure income loss is more difficult than to measure direct costs [80]. We could not find any clear patterns of methods used which affected cost estimations, except that the indirect costs in studies using reported income after diagnosis was lower than in other studies [58, 73]. Additional research is needed to validate different measurement approaches.

Where poverty impact metrics are estimated using these smaller datasets, particularly in settings where routine data collection systems are weak, there remain notable inconsistencies in their measurement.

Systematic reviews of existing patient cost studies such as the three highlighted here consistently highlight a lack of standard approaches, even where the same metric is being assessed. This can lead to challenges in assessing the comparability, quality and accuracy of results.



Guidelines for collecting patient cost data largely do not provide clear guidance on issues such as sampling, recall and disaggregation of cost data ingredients, timeframe or training of survey staff. Finally, they often provide clear guidance when data constraints require compromise.

Furthermore, reporting guidelines for economic evaluations such as the Drummond checklist largely only cover the provider perspective and don't often reflect information necessary for poverty impact metrics.

The aim this presentation is therefore to highlight some of the challenges faced in measuring the impact of illness on economic vulnerability, and spark a discussion on the importance of reporting these methodological points. To illustrate these issues, we use four case studies from our own research as examples. In doing this we hope to facilitate a discussion around how researchers can begin to report their approaches in collecting patient cost data in a more standardised and transparent way.

The Case Studies

	MERGE	XTEND	ECONPOP	REMSTART
Country	South Africa	South Africa	Burkina Faso	Zambia & Tanzania
Aim of study	Implementation and evaluation of an optimized model for scaling up TB/HIV integration at primary care clinics	Evaluation of the implementation of a new TB diagnostic, XPert MTB/RIF	Multidisciplinary study to estimate costs and consequences of abortion	Trial assessing a complex intervention to reduce mortality in ART-naive patients beginning ART
Study Design	Cluster-randomized trial	Cluster-randomized trial	Cross-sectional survey	Individually randomised control trial
Timeframe	Cross-sectional	Cohort	Cross-sectional	Longitudinal
Average cost (95% CI)	Monthly OOP expenditures: $25.82 ($16.33 - $35.33) Monthly opportunity costs of time: $43.36 ($32.64 – $54.08) Monthly income loss: $13.70 ($12.03 – $15.38)	Total OOP expenditures: $111.83 Total loan interest: $43.32 Reported income loss: $54.82 Total guardian costs: $32.11 Total carer costs: $81.99 Total episode cost: $324.07	Total OOP expenditures associated with abortion: $52.80 ($47.36 – $58.24)	OOP expenditures for one visit to study facility: $1.96 ($1.80 – $2.13)

The four case studies we present used different methods to answer very different questions. Two of the studies were conducted in South Africa and focused on TB – the XTEND study evaluated a new rapid TB diagnostic, and the MERGE study aimed to improve integration of TB and HIV services. The ECONPOP study in Burkina Faso estimated the costs and consequences of abortion, and the REMSTART study in Zambia and Tanzania implemented a complex intervention in patients beginning antiretroviral treatment.

The results from all of these studies have been published separately, the aim here is not to present or compare the study results but rather discuss a bit of the methodological decisions that went into producing them.

Which metric is used?

Metric	Lay Definition
Catastrophic overshoot	Amount by which health spending exceeds a certain proportion of income (thresholds vary between 5-40%)
Catastrophic headcount	Defined as 1 where health spending exceeds a certain proportion of income (thresholds vary between 5-40%), and 0 if not
Impoverishing expenditure (headcount)	Defined as 1 where health spending push household income below the poverty line (definitions vary between national poverty line, $1/day, $2/day) and 0 if not
Poverty depth	Amount by which health spending pushes households below the poverty line (definitions vary between national poverty line, $1/day, $2/day)

Other suggestions in the literature:

- Reflect cost-related access barriers (Moreno-Serra et al. 2011)
- More explicitly represent distributional / equity consequences of health-related spending (Ataguba et al. 2012; Onwujekwe et al. 2010, Verguet et al. 2014)
- Clarify / re-define the threshold of 'unacceptable burden' to the household (Niëns et al. 2010; Xu et al. 2003; Niëns & Brouwer 2013; Pal 2012; Wingfield et al. 2014; Onoka et al. 2011)
- Adjust expenditures to account for dis-saving or other coping strategies (Flores et al. 2008; Chuma et al. 2006; Sauerborn et al. 1996; Kruk et al. 2009)

The first fundamental methodological choice that will need to be taken by researchers is the choice of metric to use. I list here the common definitions (from Wagstaff and van Doorslaer), along with several suggestions to improve these poverty impact metrics or represent health spending in a more accurate or slightly different way. Clearly there are a large number of methods to choose from, and there is currently no authoritative guidance on which metrics to use.

With all of this choice, it's not always clear how 'catastrophic' expenditures are calculated, or how 'catastrophic' and 'impoverishing' expenditures are related. Our first recommendation is therefore just to be clear about which metric is used and how it is calculated.

Survey Comprehensiveness

	MERGE	XTEND	ECONPOP	REMSTART
Additional health services costed	Pharmacy, GP, outpatient hospital, inpatient hospital, traditional healer	Pharmacy, GP, outpatient hospital, inpatient hospital, traditional healer	None	None
Average duration of interview	~ 60 minutes	~ 45 minutes	~ 20 minutes	~ 25 minutes
OOP cost ingredients	Transport for individual & companion, medicines & consumables, diagnostics, consultation fees, special food / supplements, inpatient accommodation	Transport for individual & companion, medicines & consumables, diagnostics, consultation fees, special food / supplements, inpatient accommodation	Medicines & consumables, consultation fees, ultrasound, informal payments, pre-referral costs, hospitalisation	Transport and 'other' costs
Income data? (proxy)	Individual income before diagnosis (3-5 months prior to interview)	Annual individual income	None (GDP per capita)	Individual level income in last month

Next is the comprehensiveness of the survey. The main challenge in survey design is the representation of complex patient experiences within a manageable survey length.

Our four case studies had a range of survey durations. This is largely a function of the complexity of the patient pathways in question. MERGE and XTEND attempted to cover the overall costs of a complex illness episode over a range of different providers, whereas ECONPOP covered only a recent hospitalization and REMSTART covered only the current visit.

Disaggregation of cost ingredients will also affect survey length, and researchers may need to be strategic about the disaggregation of cost ingredients – asking in enough detail to encourage accurate recall while keeping the survey short enough that participants don't fatigue. In the MERGE study for example, we asked patients to remember the last visit to each provider – this could have been up to 5 months prior to the interview. We therefore thought it important to disaggregate cost ingredients as much as possible to facilitate recall. In contrast, as the REMSTART interview usually only covered the past few hours, we felt confident in limiting disaggregation as the potential for recall bias was much smaller.

Looking at which ingredients are included - we know that major cost drivers tend to vary by settings and even across income quintiles. For example, MERGE and XTEND which both focused on TB in different parts of South Africa found different cost drivers. The primary cost driver for the MERGE sample was special food and nutritional supplements, while for XTEND it was non-transport direct costs. This makes it difficult to highlight any particular ingredients that should or should not be included – surveys should be adapted to accurately represent what is happening in their setting. Rather we want to encourage researchers to be clear about which ingredients they do include, and how they're broken down – as the breakdown might have impacts on findings.

Finally, researchers will also need to decide how to measure income. It's generally accepted that health care spending decisions occur at the household level, and that the household is therefore the appropriate unit of measurement for patient cost surveys. However, interviews in a clinical trial are conducted individually making accurate estimation of household income difficult. In the XTEND, MERGE and REMSTART case studies, respondents consistently reported themselves to be the primary breadwinners in the household; personal income was therefore deemed an acceptable proxy for household income, with the limitation that these analyses may have underestimated the economic burden on the family as they did not account for the fact that income is shared amongst household members. With the ECONPOP sample, respondents were often not the primary breadwinners and often had no idea what the household income was. The decision was therefore made to use an assumption of GDP per capita as a proxy rather than risk breaking the confidentiality of the interview by asking family members. This decision has implications for the metrics used, as we did not have a firm understanding of where households lay in relation to the poverty line – and therefore would not have been able to report on impoverishing expenditure for example.

Timeframe

	MERGE	XTEND	ECONPOP	REMSTART
Chronic / Acute condition	Chronic	Chronic	Acute	Chronic
Study Timeframe	Cross-sectional	Cohort	Cross-sectional	Longitudinal
Recall period (costs)	The last visit to each provider (variable; max 5 months)	The last month	~ 1 day (interviewed on discharge)	< 1 day (cost of visit only)
Coping strategies considered	Borrowing, selling assets, grants or other unearned income, charitable donations, reallocation of household labour	Borrowing, selling assets, grants or other unearned income, charitable donations, reallocation of household labour	Use of entire savings, borrowing with high interest rates, selling household goods and assets, farm products, animals, reducing expenses on essential needs	None

Next, looking at timeframe – our four case studies include three chronic diseases and one acute condition.

The timeframe for the interview might in some cases be restricted by the overall study timeframe. For example, the XTEND study followed a cohort to evaluate the impact of rapid diagnostics on patient pathways, and therefore aimed to limit as much as possible the impact of the trial on those pathways. Patients enrolled in the trial could therefore only be interviewed at the end of the 6 month follow-up date. To accommodate this, the decision was made to recruit an additional sample of those on TB treatment outside the trial enrolees to increase sample size and allow for shorter recall periods between interviews.

There is also the potential for cost truncation in chronic illness or conditions with complications. The long-term economic impact of illness can be substantial. This can be a particular problem for lifelong treatments such as anti-retroviral therapy. Dissaving or other coping strategies can be an important indicator of the long-term impact of illness, and where possible it may be helpful to include coping strategies in the analysis. However, it is still relatively unclear exactly how this can be done accurately. Again, it's difficult to make recommendations at this point as to which coping strategies should be considered or how they should be incorporated into the analysis – but we want to encourage people to report their choice of methods as transparently as possible, to facilitate further research into how this can be considered.

Sample Size

	MERGE	XTEND	ECONPOP	REMSTART
Sample size	459 for costs 3478 total for trial	351 for costs 4656 total for trial	304 for economic study	1375 for costs 1999 total for trial
Subgroups (n)	TB only (n=41) TB/HIV (n=119) HIV only (n = 299)	No TB treatment (n = 302) Started on treatment (n = 49)	Induced (n=37) Spontaneous (n=267)	Intervention (n=684) Control (n=691) Tanzania (n=870) Zambia (n=505)
Catastrophic headcount (10% threshold)	44% (39% - 49%)	61% (56% - 66%)	29% (24% - 34%)	5% (4% - 7%)
Minimum sample size required to estimate proportion of catastrophic expenditure with 95% confidence	Error margin 5%: 2282 Error margin 10%: 570 Error margin 15%: 254	Error margin 5%: 1057 Error margin 10%: 264 Error margin 15%: 117	Error margin 5%: 13689 Error margin 10%: 3422 Error margin 15%: 1521	Error margin 5%: 36504 Error margin 10%: 9126 Error margin 15%: 4056

Next, coming to sample size. As we illustrate here the more infrequent the outcome of interest, the larger the sample size is required to obtain estimates within the same margin of error. Some trade-off in error margin will likely need to be made in the interests of practicality of the survey. This decision should also be taken within the context of the larger uncertainty associated with the survey – for example, spending more time in the interview to avoid recall bias may produce more reliable results than spending additional time interviewing a great many more patients.

The consideration around sample size poses particular issues for the estimation of impoverishing expenditures. As by definition patients who are already below the poverty line are not eligible to encounter 'impoverishing' expenditure, where a large number of patients are poor this outcome becomes infrequent, making power to detect the true proportion of impoverishment very low.

All three case studies estimating income had a large proportion of poor patients: 64% of XTEND patients, 45% of MERGE patients and 70% of REMSTART patients had a pre-diagnosis income below the national poverty lines. This distribution is to be expected in trials based in public facilities and investigating diseases such as HIV and TB, however this substantially reduced the power to estimate impoverishing expenditure.

Data Sources & Administration

	MERGE	XTEND	ECONPOP	REMSTART
Diary/Recall	Recall	Recall	Recall	Recall
Supplementary information collected?	None	None	None	GIS and records review through RCT
Interviewers used	Research assistants	Nurses and Research assistants	Trained female interviewers	Trained field workers
Location of interview	Facility	Facility	Facility	Facility
Medium of recording	Paper survey	Electronic survey	Paper survey	Paper survey

Finally, we come to issues around data sources and administration. This section is heavily informed by the work of DIRUM researchers in the UK – we highlight here any additional items for consideration in the context of a low-income country.

First is the choice to use diary vs. recall. Cost diaries are considered to be the gold standard in patient cost collection, but they can be time- and cost-intensive for researchers, especially where there is high illiteracy. All of our studies therefore chose to use recall, introducing some potential for recall bias.

In some cases it's possible to combat recall bias – either through retrospective records review to confirm patient visits, or through GIS mapping data to confirm reporting on travel times. The REMSTART trial had the benefit of prospective monitoring and information systems for accurate information on the number of patient visits. This trial also attempted to use GIS data to confirm reported travel times, however this proved more difficult as traffic in Dar es Salaam is unpredictable – making it very difficult to estimate an 'average' travel time for a certain distance.

There may also be a distinction in survey quality depending on the interviewer and where the interview takes place. Independent research assistants may be preferable to nurses if the subject material is sensitive; for example, the ECONPOP study asked several questions surrounding sexual and illegal behaviours which patients were unwilling to disclose to nurses. Similarly, questions on income and spending are often sensitive – and interviews can be emotionally charged in cases where a health condition pushes households into poverty. Perceived privacy will impact patient recall and willingness to disclose details on income and spending.

Finally, the medium of recording will require particular consideration in LMICs. Paper surveys can be lost or otherwise compromised resulting in missing data. Electronic data collection systems might combat these issues but do require some further training of interviewers in data entry and security.

STUDY PLANNING COMPONENT	ITEMS FOR CONSIDERATION
METRICS	• Which poverty impact metric is used? What is the chosen threshold for 'unacceptable burden' to the household, and/or how is the poverty line defined? • Are dis-saving or other coping strategies measured? If so, how are they included in the analysis?
COMPREHENSIVE-NESS OF SURVEY DESIGN	• Which OOP expenditures are included? What is the level of disaggregation in cost ingredients and how long is the survey? Are any context-specific variables included? • How is income measured, and for what timeframe (i.e. current vs. pre-diagnosis)? Whose income is collected (i.e. personal or household income)?
TIMEFRAME	• What is the timeframe of the survey? Is it appropriate to capture all economic outcomes? • What is the complexity of the disease pathway? Is there resulting potential for recall bias? Is there potential for cost truncation in the context of chronic disease and/or future complications? • Are coping strategies used to estimate the long-term economic impact of health spending?
SAMPLE SIZE	• What is the confidence interval and margin of error deemed acceptable? • If estimating impoverishing expenditures, what is the distribution of pre-diagnosis income below the poverty line? Are any adjustments to sample size required to account for clustering, or non-response?
DATA SOURCE & ADMINISTRATION	• Is a cost diary or recall used to capture expenditures? • Is data supplemented with any additional data sources, such as retrospective records review or GIS data? What is the medium of collecting and recording data (i.e. electronic, paper, or telephone surveys)

Our recommendations for reporting methods are listed here in a summary table.

This session confirms the increasing implementation and sophistication of economic evaluation in LMICs. Going forward in these settings, evaluations need to tackle policy concerns around equity and poverty. We are seeing that poverty impact metrics are increasingly important to policy makers and programme planners, however they are currently data-hungry and inevitably there will need to be some degree of compromise in the planning stages of a project due to time- and budgetary constraints in a clinical trial.

There is currently little understanding about where compromise may be acceptable. We therefore advocate for further methodological work is to investigate the means to minimize the impact of compromise when planning poverty impact studies. We also encourage researchers to report data sources as transparently as possible, both to facilitate this further methodological work and to help guide each other going forward in collecting this data.

Thank you!

We are grateful to the study teams for the MERGE, XTEND, REMSTART and ECONPOP trials for use of their data and for their insight on challenges encountered in data collection.

The **MERGE** study team includes Tendesayi Kufa, Piotr Hippner, Salome Charalambous, Katherine Fielding, Alison Grant, and Gavin Churchyard. The **ECONPOP** study team includes Johanne Sundby, Gaute Torsvik, Katerini Storeng, Veronique Filippi, Fatoumata Ouattara, Hanne Lichtwark, Ramatou Ouedraogo and Seydou Drabo. The **XTEND** study team includes Susan Cleary, Lucy Cunnama, Gavin Churchyard and Edina Sinanovic. The **REMSTART** study team includes Sode Matiku, Bernard Ngowi, Duncan Chanda, Sokoine Lesikari, Christian Bottomley, Saidi Egwaga, Amos Kahwa, Peter Mwaba, Sayoki Mfinanga, and Shabbar Jaffar.

We are also grateful to Catherine Pitt and Ulla Griffiths for their comments on an early draft of this presentation

same
Modelling & Economics
http://blogs.lshtm.ac.uk/samegroup/

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

This presentation is the result of four large trials and could not have been done without the insight gained from each. Our enormous thanks to everybody involved in all of the studies.