Author's response to reviews

Title: A framework for the improved use of routine health system data to evaluate national malaria control programs: evidence from Zambia

Authors:

Adam Bennett (abennett@tulane.edu)
Joshua Yukich (jyukich@tulane.edu)
John M Miller (jimiller@path.org)
Penelope Vounatsou (penelope.vounatsou@unibas.ch)
Busiku Hamainza (bossbusk@gmail.com)
Mercy M Ingwe (nmercie@yahoo.com)
Hawela B Moonga (mhawela@yahoo.co.uk)
Mulakwo Kamuliwo (mkamuliwo@yahoo.co.uk)
Joseph Keating (jkeating@tulane.edu)
Thomas A Smith (Thomas-A.Smith@unibas.ch)
Richard W Steketee (rsteketee@path.org)
Thomas P Eisele (teisele@tulane.edu)

Version: 2 Date: 18 April 2014

Author's response to reviews: see over
April 17, 2014

Editors, Population Health Metrics

Dear Editors,

Thank you for your careful review of our manuscript titled ‘A framework for the improved use of routine health system data to evaluate national malaria control programs: evidence from Zambia’ (Manuscript ID 1092321629110604). Enclosed is a revised manuscript in both track changes and in a clean format, as well as the revised Additional File (clean and in track changes). We would be grateful if this manuscript could be reconsidered for inclusion in Population Health Metrics. We have included a point-by-point response to the issues raised by the 3 reviews. We feel the revised manuscript has been strengthened considerably based on the thoughtful comments by the reviewers and editor.

We hope this revised paper now acceptable for inclusion in the Journal.

Sincerely,

Adam Bennett, PhD
Assistant Adjunct Professor
**Point-by-point response to reviewers**

**Reviewer #1**

The questions the authors are trying to answer are of first order importance in malaria control monitoring in Sub-Saharan Africa where routine data systems are fraught with significant weaknesses. The study uses a new framework that offers promise on how to adjust routine data for program monitoring and evaluation. I think there would be significant wider policy interest in this type of work. I read this paper with great interest.

That said, I have a few issues with the paper:

R1C1: I am quite concerned that the authors do not pay sufficient attention to well-known data quality problems with the HMIS, aside from completeness. For example, number of cases of (confirmed and unconfirmed) malaria in the HMIS at least at facility level are quite noisy, with huge, unexplained spikes over short time periods. Did the authors check for implausible numbers of reported malaria cases? Further, some small studies have shown that even what is recorded as confirmed malaria on the HMIS is not actually confirmed with RDT or microscopy. Are the authors sure that “In cases where a confirmed case count was reported but no parasitological testing value reported, we replaced the missing testing value with the number of confirmed cases” does not raise any questions whether what are classified as confirmed were indeed tested? How many are such cases, and are they mostly in children or in certain parts of the country?

Finally on this issue, I would have also liked to see the all-cause morbidity patterns shown in Figure 4 or in separate Figure altogether.

We thank the reviewer for highlighting these concerns. We did not initially remove outliers from the dataset. As a check for potentially spurious values, we visually inspected individual health facility counts and then examined dispersion from the median using simple outlier statistics such as the median absolute deviation. While in many cases these data are noisy at the facility level, we did not identify any “implausible” values.

We also conducted validity checks to see if our predictions were biased. Accordingly, we have added the following description to the Additional File: “We ran validity checks on this prediction method by withholding a subset of the data (15%) and did not identify any systematic bias in predictions by facility type, district, or month. We found a mean prediction error of -3.7 for health centers, -1.6 for hospitals, and -0.81 for health posts. Overall the mean prediction error was only -3.4, indicating very little bias. Total prediction error was roughly 9%, which is very slight especially as we only imputed 21% of total case values and 37% of confirmed case values.”

We also ran the models with potential outliers removed, and found very similar effect estimates. We have added the following description of this exercise to the Additional File: “Additionally, we ran a sensitivity analysis whereby we ran the initial imputation model with all available data, and ranked the percentage error residuals from predictions from this model. We then removed the bottom 2.5% and top 2.5% from these residuals (potentially low and high implausible values based upon the entire dataset). We then reran the models without these outliers and obtained similar effect estimates, thereby indicating that potential outliers were not driving our findings.”
With regards to testing value, we cannot be certain that confirmed cases without a testing value were indeed confirmed. Roughly 33% of lab testing values were replaced with the confirmed case value. We have added this description to the second paragraph on page 7, to the sentence that now reads: “In cases where a confirmed case count was reported but no parasitological testing value reported (roughly 33% of all testing values).”

That said, we believe this discrepancy is most likely to be almost entirely due to reporting of testing, as this was not a standard reporting item before 2009. The largest number of missing lab testing values (where there were corresponding confirmed case values) occurred early in the reporting change, in 2009.

We also include this as a limitation in the discussion paragraph 6: “While we attempted to control for the increase in confirmed case testing in multivariable models, our testing rate may be an imperfect indicator of the true testing rate, as reporting of testing likely improved contemporaneous with RDT scale-up, laboratory testing values were not consistently reported, and detailed RDT stock-out data were not available.”

We found no distinct differences by age group. The highest proportion occurred in Northwestern and Copperbelt provinces, and the lowest occurred in Central and Lusaka provinces. Finally, excluding the testing rate term from regression models did not greatly alter our results.

We thank the reviewer for the suggested change to figure 4 - we have added total OPD to the figure as requested.

R1C2: The authors mention that for districts and/or months with missing data they had to impute. But they do not tell us what proportion of the data were ultimately imputed, and what the possible implications of removing those districts would have been on the analysis.

We imputed at the facility-month level rather than the district level, we have clarified the following statement in the methods on page 6: “Before aggregating to the district level, we imputed all missing facility-level monthly outpatient malaria values based upon the spatial location of the facility and the month in which it occurred using Bayesian conditional autoregressive models” We have also added the following clarification to the first paragraph of the results: “Of the 60,948 possible facility-month observations, there were 48,166 (79.0%) non-missing values available for total malaria cases, and 38,588 (63.3%) non-missing values for confirmed cases alone; the remaining 21.0% of total cases and 36.7% of confirmed case values were imputed.”

Additionally, we ran the model without these data imputed (as missing) and obtained larger estimates of ITN effectiveness, indicating that our imputation is likely a more conservative approach in this setting.

R1C3: The authors need to be more transparent about the methods used to estimate district level ITN coverage. What set of covariates was included in the model used, if any? Also, how good are the predicted district-level coverage rates?

We have improved the description of the ITN coverage estimation in the Additional File, including description of covariate inclusion and the validation exercise we conducted for the geostatistical
predictions as follows: “To produce ITN coverage estimates for each district over the study period the following steps were completed to combine available household survey and enumeration data and program distribution data. In step one, data were first compiled at the cluster level from the 2008 and 2010 MIS surveys, district-level surveys conducted in Luangwa and Nyimba districts, and program household enumeration data on the total number of ITNs and the total number of household members per cluster. There were 400 clusters included for 2008, and 245 clusters for 2010. The ratio of ITNs to persons at each cluster was then modeled in a Bayesian geostatistical framework with a normal (Gaussian) prior and with covariates urban/rural and distance to the district health office. An exponential spatial decay parameter was included to capture spatial autocorrelation between clusters and allow creation of a spatial prediction surface. All models were fit using WinBUGS. Model convergence was assessed using plots of ergodic averages after a burn-in period of 5,000 iterations. For the geostatistical models, predictions to each 5km x 5km grid cell covering Zambia (a total of 36,159 grid cells) were produced using the spatial.unipred command (Figures AF 1-2). We conducted a validation exercise whereby we withheld a 15% subset of data for prediction, and ran each model on the remaining 85% training dataset. We found consistent slight under-predictions (mean prediction error of -0.10 in 2008 and -0.05 in 2010) from these validation models, but prediction errors were not spatially patterned. 61% of true values fell within the 95% Bayesian Credible Interval (BCI) in 2008, and 83% of true values fell within the 95% BCI in 2010.

We calculated population-adjusted district means from the predictions of these models by multiplying each ITN per person surface with a Landscan population raster\(^6\) adjusted for annual population change. In step two, we used program ITN distribution data and district populations for each year to separately predict the potential availability of ITNs per person per district from distributions over the study period by applying a decay function based upon the NetCalc algorithm to the previous three years’ distribution counts.\(^7,8\) This algorithm assumed a three-year half-life for ITN decay. Finally, in step three we built a linear regression model with random intercepts for each district to assess the relationship between the predicted availability from distribution data and the modeled coverage estimates for 2008 and 2010 from step one, and used this model to interpolate ITN per person estimates for 2009 and 2011 (Figure AF 3). Model assessment revealed that incorporation of spatial effects in step three did not improve model fit and therefore only uncorrelated random intercepts were included in this final interpolation step.\(^9\)

For initial geostatistical models we included distance to the district health management office (DHMO) and rural/urban as a covariate, based upon the known potential for differential access to nets based upon these variables (Noor et al AJTMH 2010).

R1C4: I sympathise with the fact that the paper was no focused on examining the lack of an expected association between ITN coverage in high transmission intensity. Could this issue be connected with testing and reporting behavior? Further, we know that distribution of ITN is not exogenous but endogenously related to malaria incidence? I would like to hear the authors’ opinion on this since these issues arise naturally in the reader’s mind.

Indeed we have bolstered the discussion of endogeneity in the discussion in the second to last paragraph, as follows:
Second, potentially endogenous relationships existed between our primary outcomes and explanatory variables of interest due to programmatic choices targeting high burden or easily accessible areas. In some instances, such as the use of calendar month in the evaluation of IRS effectiveness by Over and colleagues,\textsuperscript{38} instrumental variables may be available to infer causal relationships when endogeneity exists. However, as no instrumental variables uncorrelated with primary outcomes were available in our data, we were not able to perform two-stage regression or similar standard econometric approaches to isolate uncorrelated effects. Rather, we controlled for systematic spatial targeting of intervention effort through the use of anomalies in program coverage. This approach was effective for ITN coverage, as the goal is for universal coverage and therefore targeting has been limited. However, the highly targeted nature, relatively far lower coverage of the IRS program during this period, and our lack of confirmed case data preceding IRS scale-up precluded our ability to make similar effective adjustments for IRS. Future use of these data will likely prove more robust for evaluating IRS efforts as more areas are included and there is greater heterogeneity within districts over time.

We have also added the limitation that testing and reporting may be influencing the relationship between high burden and ITN coverage in the fourth paragraph as follows: “It is also possible that our testing and reporting rates do not fully correct for biases in diagnostic reporting practices in these high burden areas.”

R1C5: The period of the analysis coincides with removal of user charges in public health clinics in rural Zambia (from 2006) and urban health clinics (from mid 2011). This policy has been shown to have increased facility utilization rates. Further, a number of districts have relied heavily on volunteers community health workers to administer ACTs and even testing. These cases do not enter the HMIS, what are the possible implications for these findings?

We thank the reviewer for this comment, and acknowledge that this is a possible limitation. However, the removal of user charges (2006) well predates the beginning of our study period (2009) and therefore was already in place at the time of our study and therefore likely few changes from 2009-2011 can be attributed to this policy. That said, it is possible this may have affected some urban facilities at the end of our study period in 2011. There was some evidence of increasing all cause OPD over this period in some areas (from Figure 4), but it is difficult to ascertain whether this increase caused increases in total malaria OPD, or was caused by increases in total malaria OPD, given the large proportion of all OPD due to malaria. By comparison, increases in non-malaria OPD were slight. We have added the following sentence to the fourth paragraph of the discussion: “User fee changes adopted in 2006 may have influenced health facility utilization rates broadly, but the bulk of these effects would likely have well predated our study.”

Furthermore, there was little evidence of increasing utilization in household surveys over this period.

R1C6: Shortages of RDTs often influence clinician’s classifications of malaria as confirmed or unconfirmed. The authors do not discuss the implications of this on their findings.

We agree with the reviewer’s assessment. Unfortunately, detailed RDT stock-out data were not available nationwide for this period. We have added the following to the limitations: “...laboratory testing values were not consistently reported, and detailed RDT stock-out data were not available.”
MINOR
R1C7: What do the authors think about interacting testing rates with transmission intensity?

We thank the reviewer for this suggestion. However we do not include separate interactions between testing rates and transmission (either prevalence or incidence) as, while we were interested in describing changes in testing rates, in regression models we included testing as a control variable only. Further, the testing rate is highly correlated with one of our possible transmission indicators, confirmed case incidence. The inverse relationship between the testing rate and total malaria OPD found in the regression model signifies that indeed the testing rate appears lower in higher burden areas.

R1C8: In terms of placing the observed magnitude of effect on malaria incidence into the relevant literature, it would be very informative if the authors examined net usage instead of or in addition to, net ownership as, after all, is a key mechanism for the health impact.

We did not examine net use as part of this study as, in order to improve the precision of ITN coverage estimates, we included data from enumerations conducted in several districts. These enumerations captured only ITN ownership, not use. However, we modeled the ITN density ratio (ITNs per household member) rather than household ownership of at least one ITN. Previous studies have shown that this measure of ITN access correlates well with use (Bennett et al Plos One 2012; Garley et al Malaria J 2013).

R1C9: Is there a reason why we should expect the standardized reporting rate to have a positive effect on confirmed malaria but not on total malaria? What is the authors’ intuition on this finding?

We state in the results that “The standardized testing rate was positively associated with confirmed malaria case incidence, but was negatively associated with total malaria case incidence, whereas the standardized reporting rate was positively associated with total malaria case incidence, but not with confirmed case incidence.” It is likely that increased confirmed malaria is due to increased testing over time, and testing was increased simultaneously with reporting. Similarly, this seems to suggest testing rates are higher in lower total malaria burden areas. It is likely that the take-up of RDT testing, and reporting of that testing, is poorer and less accurate in high burden areas. Moreover there is a possibility that testing is less needed in higher burden facilities as presumptive diagnosis will be more specific there. By contrast, reporting rates were not associated with confirmed case malaria, but were higher in high burden areas.
Reviewer #2

The study by Dr Bennett and colleagues used the Zambia HMIS for 2009-2011 and explanatory information from a number of household surveys and programmatic data to assess the associations of ITN coverage with changing malaria incidence in Zambia. The study shows that ownership of 1 ITN per household was associated with overall 27% reduction in clinical incidence. The Framework developed by the authors is a useful formal approach towards assessing the impact of malaria interventions on disease burden. The approach is easily adopted for other countries in Africa where the reporting rate is reasonably high and where sufficient data on intervention coverage exists. My comments are as follows:

R2C1: It is not clear to me why reporting rate and testing rates are included as dummy variables in the regression if they were used in the process of imputation of missing case data. Were they used in the imputation?

We did not use the reporting rate and testing rates in the imputations. We state the following regarding the imputation under “Primary Outcomes”: “Before aggregating to the district level, we imputed all missing facility-level monthly outpatient malaria values based upon the spatial location of the facility and the month in which it occurred using Bayesian conditional autoregressive models. “

The imputations were conducted at the health facility level, whereas the reporting and testing rates are aggregates entered into the regression framework at the district-month level. We state the following regarding the reporting and testing rate: “To evaluate reporting rates over time we created an index of the number of facilities reporting per district per month as a proportion of the total number of facilities reporting per district, weighted by facility size (determined by mean monthly malaria outpatient diagnoses over the study period). We created a similar index for testing per district-month calculated as the total number of parasitological”

R2C2: It is not clear also whether the authors have adjusted the clinical cases using the test positivity rate (TPR) at each health facility so that their overall case load per facility per month is = confirmed cases + (unconfirmed cases adjusted for TPR of the month). If this was not done, then chances are that where TPR is low, the authors will primarily be looking at the effect of ITN on fever burden as opposed to malaria burden.

Indeed this is an important consideration. However, we did not adjust the clinical cases by the test positivity rate. We state that: “The primary outcomes included monthly confirmed and total (confirmed + unconfirmed) outpatient malaria cases aggregated at the district level.” In many cases, testing information is not available at the facility level. Further, it is likely that the overall (confirmed + clinical) incorporate some level of non-malarial fever burden. This highlights the need to present analyses on both overall case rates and confirmed case rates, as we do.

R2C3: The authors adjust for treatment seeking rate but this usually a measure of access and not necessarily use of effective medication. I think the proportion of fevers treated with appropriate antimalarial drugs is equally important. I am aware that harmonizing this variable across clinically vs parasitologically
diagnosed cases is tricky without information from surveys on whether febrile cases were tested for malaria before treatment.

While the proportion of fevers treated is an important indicator for ACT coverage, we do not investigate treatment in this analysis. We incorporate only treatment seeking as a potential confounder, as differential treatment seeking rates may influence facility attendance, and thereby outpatient case levels. This is true irrespective of treatment. As the reviewer suggests, these are can be tricky to harmonize given the lack of reliable information on diagnostics in household surveys, however this was not the goal of this analysis.

R2C4: The authors have used the MAP 2010 estimates of PfPR2-10 to classify the country by malaria endemicity class. They also found interaction between ITN coverage and these endemicity classification. I think there are a number of issues when using these 2010 estimates to determine effect of ITN coverage. The transmission maps reflect the effect of intervention coverage up to 2010 in Zambia. However because the HMIS data used starts from the year 2009 the impact of these problem on the models may not be great. Of more importance is that in developing the 2010 MAP products a climatic covariates similar to those used in the regression analysis were used. This creates a potential problem of circularity. While we did examine interactions with the MAP 2010 surface as a proxy for transmission, we did not ultimately use the MAP interaction, in part due to the concern suggested by the reviewer. The interaction we used in models was between ITN coverage and the three highest incidence burden provinces. We state the following in the methods: “Exploratory and residual analysis revealed potential interactions by region between primary outcome and explanatory variables. In model construction we therefore assessed the inclusion of interactions between ITN coverage and transmission, as measured by PfPR2-10 (Malaria Atlas Project) categories (<10% vs. >10% and <25% vs. >25%), as well as between ITN coverage and high burden/low burden province, where high burden provinces were those with the highest confirmed case incidence over the entire period (Luapula, Copperbelt, and Eastern provinces as defined in 2011).”

We have included the following in the fourth paragraph of the discussion (previously this was in the Additional File): “Our finding of a significant interaction between ITN coverage and low versus high incidence regions in models predicting both confirmed and all malaria outpatient cases was unexpected; while potentially related to transmission, we did not find significant interactions between district ITN coverage and endemicity categories as defined by mean PfPR2-10.”

R2C5: The authors could have modeled ITN use in the same way they modeled ITN ownership as the former has a direct relationship with observed changes in incidence. Any reason why this was not done?

We chose not to model ITN use as we included enumeration data to increase precision of ITN estimates, and use information is not available in those data (see previous). Instead we use ITN access (ITNs per person), which is known to correlate well with use (Bennett et al Plos One 2012; Garley et al Malaria J 2013).
Reviewer #3

I enjoyed reading this very interesting paper. The methods represent a significant advance on previous similar work. The use of both confirmed and clinical malaria cases, the use of anomalies to adjust for seasonal variation and the inclusion of climate variables was commendable. The problem is that there are relatively few years analysed (2009-2011) and these were years of very significant changes in diagnostic availability reporting and other factors, as the authors acknowledge. The lack of programmatic data on net distribution led to need for extrapolation from surveys and the inability to adjust for net distribution or durability over time (by month). The multiple layers of assumptions on top of spatial predictions make the methodology hard to assess clearly, especially for a reader not fully up to date with Bayesian techniques.

I think the short Results section in the abstract about one small and (in my opinion rather misleading) association with nets (since most of it seems to be coming from the low endemic areas) did not capture at all the main points in the paper. My reaction on reading it was

1. this paper introduces some significant methodological improvements to the assessment of malaria control effectiveness, most notably the inclusion of diagnostic and reporting rates and methods to extrapolate spatially from cross-sectional surveys to net coverage indicators by district (although one would not need to do that if one had the programmatic data)

2. malaria is increasing in Zambia in most regions, and nets do not seem to be having impact on that in the highest incidence areas

3. that control programmes would have a hard time repeating this analysis on a regular basis. We need to work towards simpler and clearer methods that can be applied in country by the control programmes.

I think this paper would be better split into two (although I realize that has cost implications in this day and age). One could be the methodological approaches and perhaps description of the malaria situation and interventions/changes applied, and the other the application of the methods to the Zambia data to draw conclusions about effectiveness of different control methods. Currently, the interpretation of the data and its summary in the results section gets the least attention given all the effort needed to describe the methodological issues. The most important concept that control programmes do not seem to usually grasp is the necessity to quantify inputs (nets, spraying etc) in the same geographical and time units as the outcomes (incidence by district-month, for example). This paper could contribute to that process but that seems to be a bit lost in here. If (e.g.) net distribution could be tracked by district and month, it would not be necessary to do the complicated extrapolation from relatively crude ITN cross-sectional survey data.

A lot of effort has been gone to here to gather complete data for every district – however while it is valuable to have data everywhere for monitoring of the
program, more sophisticated analyses to quantify effect size for particular interventions can be done using a subset of districts that have complete data. Another reason to split it into two.

R3C1: The abstract refers to ‘district-time’ units in the Methods. I initially assumed that the time unit was month for all variables but became a bit confused with the net data when I realised it was annual (‘Methods: Measures of primary exposure variables’) and apparently included as a constant in each year of the study? Could it not have been adjusted or interpolated somehow over time? As it stands, and given all the assumptions made, I don’t find the results on impact of nets too convincing.
Does Zambia not have the records of when all the nets were distributed, and where?
We thank the reviewer for highlighting this limitation. We considered methods to produce more temporally detailed net coverage estimates, and monthly records are available on when the central program distributed nets to the districts. However, we know through experience in Zambia and anecdotally that while these data approximate shipment times to districts, they are by no means accurate indicators of the timing of distribution to households, and furthermore these data have additional quality issues. We feel in this setting it is more valid to center coverage estimates on measured household survey data where available.
We examined linearly interpolating this annual value so that we produced a varied monthly value for each district – and this produces an even larger effect estimate for ITNs on incidence. We chose not to use this method as it may inject additional bias if ITNs were not scaled up consistently throughout the year. Indeed, in most years distributions occur preceding the next years rainy season (late in the calendar year to precede the rains that begin in late November). We have added the following statements to the discussion to highlight this limitation, second to last paragraph: “Additionally, we incorporated only annual ITN coverage data, which may not accurately depict monthly changes in coverage. There is need for programs to more closely track monthly ITN coverage data in order to make more temporally-refined assessments of intervention effectiveness.”

R3C2: in the Methods: Measures of primary exposure variables para 2, the IRS data is dismissed rather quickly as being endogenous and not considered further (although included in the model as ‘control variable’). Endogeneity is definitely likely, but we really need some more basic information on how much IRS was done, where and when, given the focus on IRS in Zambia and the issue of whether to do combined IRS and ITN in the global community these days. It seems glib to just dismiss it as ‘oh we can’t tell what is happening due to IRS’ since that is the whole point of studies such as this, to determine the effectiveness of all interventions. More exploration of the endogeneity issue and how to deal with it would be welcome, since IRS impact has to be assessed somehow or why are we doing it? There’s a bit too much emphasis on ITN. Did you look at calendar month as instrument as in Over et al 2006 AJTMH?
We thank the reviewer for highlighting this important concern. Indeed we considered the approach by Over, but we do not think it will work with our data to isolate effects of IRS, for two primary reasons: 1) IRS is conducted very focally both spatially and temporally in Zambia, often within only 1 or 2 months of
the year preceding the wet season, and therefore there are almost no data in other months, and 2) calendar month did not function well as an instrument for IRS exposure, as month remained highly correlated with the outcome (incidence of confirmed malaria cases) even after controlling for rainfall, vegetation, and maximum and minimum temperature. We have added the following section to address this in the limitations in the discussion: “In some instances, such as the use of calendar month in the evaluation of IRS effectiveness by Over and colleagues,38 instrumental variables may be available to infer causal relationships when endogeneity exists. However, as no instrumental variables uncorrelated with primary outcomes were available in our data, we were not able to perform two-stage regression or similar standard econometric approaches to isolate uncorrelated effects. Rather, we controlled for systematic spatial targeting of intervention effort through the use of anomalies in program coverage. This approach was effective for ITN coverage, as the goal is for universal coverage and therefore targeting has been limited. However, the highly targeted nature, relatively far lower coverage of the IRS program during this period, and our lack of confirmed case data preceding IRS scale-up precluded our ability to make similar effective adjustments for IRS. Future use of these data will likely prove more robust for evaluating IRS efforts as more areas are included and there is greater heterogeneity within districts over time.”

R3C3: Intro para 2 and 3 – I think the description of previous longitudinal time-series studies as ‘simple’ is not really valid (many used climate variables and/or multiple sites for example e.g. Thomson et al in Botswana, Coetzee in S Africa, Loha and Lindtjorn in Ethiopia 2010, Abeku studies in Ethiopia) and the novelty of the district-time approach is slightly overstated. Apart from being a common econometric approach, not just invented by Victora for health, it was applied for malaria in both Over et al and the Eritrea studies Graves et al (cited in discussion). It’s true previous studies did not adjust for everything mentioned: diagnostics, reporting etc but did have better data on monthly intervention inputs and used imputation for the reporting issues. Please revise these paras.

We thank the reviewer for highlighting these concerns. In our intro we chose not refer to longitudinal cohort studies (as Loha and Lindtjorn) or other studies mentioned that do use HMIS-type routine data to examine climate effects as our focus is primarily on the use of these data for program evaluation. We have clarified this in the intro with the following changes. In the second paragraph, we have revised to: “Although time series HMIS data have been used for sophisticated climate modeling and early warning systems,4 to date most uses of HMIS data for program evaluation in Africa have been simple comparisons of pre and post-intervention trends in rates of malaria case incidence and deaths.5 Only in rare cases have such studies directly controlled for important confounding factors, including changing diagnostic confirmation practices, access and use of health services, HMIS completeness, and rainfall”. We have added the following clarification to the third paragraph: “Graves and colleagues (2008) previously used such an approach in their evaluation of vector control scale-up in Eritrea on the outcome of HMIS-derived malaria case incidence, while accounting for climate variability.8 However, while their study is a significant advancement over simple analysis of HMIS trends over time, they did not account for malaria diagnosis practices, health services access, treatment seeking and spatial and other unobserved correlations in the data.”

R3C4: Please be consistent in use of ITN ‘coverage’ terms (coverage includes ownership, use and access) and distinguish the definitions. On page 4 we have “ITN household possession” (= % of HH with at least one?), on page 5 we have ‘ITN program intensity’, page 11 we have ‘ITN:household ratio’ para 2, and
‘overall ITN coverage’ in para 3, Figure AF3 shows ITN per person’ etc – please clarify and be very consistent about the variables used in models and elsewhere. Indeed we aimed to clarify the use of these related terms throughout. We have clarified in the methods as follows, under “Measures of primary exposure variables”:

“The primary exposure variable for this analysis was ITN coverage measured as ITNs per household at the district level per year. Bayesian geostatistical models were first used to produce estimates of ITN per person ratios from National Malaria Indicator Surveys (MIS) and IRS program enumeration efforts in 2008 and 2010,9,13 and population-adjusted values calculated per district (see Additional File Figures AF 1 and AF 2). Bayesian generalized linear models were then used to predict values of ITN per person ratios for districts and years without survey data from annual district ITN distribution data from the National Malaria Control Center (NMCC) (Figure AF 3). The resultant district-level ITN per person ratio was multiplied by the average household size of each district in order to represent population coverage as a more programmatically useful value, the number of ITNs per household. In final regression models, we included this number of ITNs per household variable as an anomaly from the four-year mean for each district to control for systematic spatial effects and potentially endogenous relationships due to programmatic targeting decisions.”

R3C5: On page 11 we suddenly get a reference to AF 6 – what do the other parts of the additional files add? Please include some reference to them, put in main doc or leave them out.

We have added earlier references to the Additional file figures.

R3C6: Please cite additional files in order and you have two AF 6 figs, please resolve

We have made these changes.

R3C7: Discussion para 2, last sentence, it is not quite true that studies did not try to account for correlated data on malaria over time – we did include malaria cases lagged one month in the Eritrea studies to try and get at the time correlation, so please note that. For spatial correlation, that is true.

We have edited paragraph 2 of the discussion to read: “Finally, none of these studies accounted for the inherent correlated nature of malaria case data across spatial units, which can result in erroneous findings of statistical significance if not accounted for, and only the Graves study accounted for temporal autocorrelation.”

R3C8: It is unfortunate that inpatient cases were not also considered, as they may be less subject to the diagnostic and reporting biases of outpatient cases. It is stated in the ‘Methods- Study site’ that these were available. Please either include them or comment on why inpatient cases and deaths were not used (I assume for deaths there were too few but please comment).

We chose not to include inpatient cases and deaths as, while it is reasonable to assume these cases will be less likely to be incorrectly diagnosed, we found similar rates of reporting and testing (and therefore potentially similar bias). As the reviewer suggests, there were indeed few deaths in many parts of the country. Additionally, the use of inpatient cases is potentially influenced to a greater degree by other facility-level factors. Finally, inpatient cases are more likely to be potentially influenced by confounding factors of population immunity.
As mentioned above, please consider splitting into two papers with one focused on methodology and one on the results of the analysis of effectiveness of control methods. (using inpatients as well if possible)

We have additionally added the following to the results section reflecting this reviewer’s concerns:
“Total reported malaria outpatient cases increased as the HMIS reporting system strengthened, from 3.0 million in 2009, to 4.1 million in 2010 and 4.3 million in 2011; this corresponded to an increasing annual parasite index (API) per 1000 population of 99.8 in 2009, 135.9 in 2010, and 194.8 in 2011 using only confirmed cases and imputing missing values. After controlling for confounding factors, an increase...”

Editorial comments
We would be grateful if you could address the comments in a revised manuscript and provide a cover letter giving a point-by-point response to the concerns. The manuscript’s handling editor would like to note that you do not need to follow the suggestion by a reviewer to split this into two papers. Please address in detail all the comments on the methodology and data in the reviews by Masiye and Noor. In addition, please also provide more details on:

The imputation of HMIS data (including how much data were missing, how many facilities do you have data for, how representative is the sample of the whole country, especially given that not all facilities, including some large facilities do not report to the HMIS). Did you do any validity checks of the method to see if it is an appropriate method to use in this dataset? Why was this method selected? Were alternative methods considered? In the supplemental file, you say that you estimated catchment area for each facility based on the number of outpatient cases in 2011? how did you do this and how can you justify using outpatient cases to estimate catchment area (this method assumes that the probability of seeking care is evenly distributed in the population? which is highly unlikely to be true)?

We have included the following information to the Additional File covering our validation and sensitivity analyses:
“We ran validity checks on this prediction method by withholding a subset of the data (15%) and did not identify any systematic bias in predictions by facility type, district, or month. We found a mean prediction error of -3.7 for health centers, -1.6 for hospitals, and -0.81 for health posts. Overall the mean prediction error was only -3.4, indicating very little bias. Total prediction error was roughly 9%, which is very slight especially as we only needed to impute 21% of all values.”

“Additionally, we ran a sensitivity analysis whereby we ran the initial imputation model with all available data, and ranked the percentage error residuals from predictions from this model. We then removed the bottom 2.5% and top 2.5% from these residuals (potentially low and high implausible values based upon the entire dataset). We then reran the models without these outliers and obtained similar effect estimates, thereby indicating that potential outliers were not driving our findings.”

With regards to facility catchment area, we have updated the Additional File text to read as follows:
“We then imputed missing values using conditional autoregressive models (CAR) with existing data values, the spatial neighboring relationship, a first-order temporal autoregressive term within WinBUGS, and with the estimated facility catchment size as the exposure. As we lacked facility catchment size information for most facilities, catchment sizes were estimated from the monthly average outpatient attendance for all diagnoses in 2011. Standardizing the malaria case data in this way increased the spatial correlation in the data and thereby increased the information available for imputation."

A similar approach was used by Gething and colleagues (Plos Med 2006) in their imputation of HMIS data in Kenya (cited here).

· The estimation of ITN coverage rates: the MIS does not sample in all districts, and it’s not meant to be representative of the district the cluster is located in. Why did you not use data from the DHS on ITN coverage as well? The methods are not clear: a lot more detail needs to be provided on how ITN estimates were derived to be able to assess the validity of the methods.

We detail the methods for estimating district level ITN coverage in the Additional File. For initial geostatistical models we included distance to the district health management office (DHMO) and rural/urban as a covariate, based upon the known potential for differential access to nets based upon these variables.

Similar to the imputation exercise, we ran a validation based upon a 15% subset of data for each geostatistical surface. We found consistent slight under-predictions (mean prediction error of -0.10 in 2008 and -0.05 in 2010) from these models, but prediction errors were not spatially patterned. Additionally, as we use anomalies as our exposure term, this is unlikely to have injected bias. We have added the following statement to the Additional File: “We conducted a validation exercise whereby we withheld a 15% subset of data for prediction, and ran each model on the remaining 85% training dataset. We found consistent slight under-predictions (mean prediction error of -0.10 in 2008 and -0.05 in 2010) from these validation models, but prediction errors were not spatially patterned. 61% of true values fell within the 95% Bayesian Credible Interval (BCI) in 2008, and 83% of true values fell within the 95% BCI in 2010.”

· The estimation of fever rates per district: since the surveys you mention are not representative at the district level, how did you produce district-level estimates?

While each survey is not mean to be representative at the district level, we combine data from six surveys covering 2006-2011. We examined simple kriging models based on these data, and additionally examined trends in fever seeking over time. We found consistent rates of fever treatment seeking over time, and found no difference when using a kriging method to interpolate treatment seeking. We have added the following statement to the methods: “We examined simple kriging methods but found no difference with these cross-survey district summaries.”

· Figure AF6: how do you explain negative numbers of ITN per HH on the graphs? Were the coverage estimates not constrained to be in the range [0,1]?

As we state in the methods, we use anomalies from the district mean, so negative numbers refer to decreases in ITN per HH from that district mean ITN per HH over the study period.

Please use track changes or colored/highlighted text to show your revisions. Please also ensure that your
revised manuscript conforms to the journal style (http://www.pophealthmetrics.com/info/instructions/). It is important that your files are correctly formatted.