PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Measuring effective coverage of curative child health services in rural Burkina Faso: a cross-sectional study
AUTHORS	KOULIDIATI, Jean-Louis; Nesbitt, Robin; Ouedraogo, Nobila; Hien, Hervé; Robyn, Paul; Compaore, Philippe; Souares, Aurélia; Brenner, Stephan

VERSION 1 – REVIEW

REVIEWER	Katya Galactionova Swiss Tropical and Public Health Institute Switzerland
REVIEW RETURNED	22-Aug-2017

GENERAL COMMENTS	Frameworks for evaluation of health service coverage are flexible and can be used to capture a broad spectrum of service aspects; differences in indicators adopted by researchers to evaluate health interventions enrich the field with new insight on service delivery. The authors of this study draw on rich data covering both provider and patient side factors and present new evidence on the quality of child health services in Burkina Faso. The analysis and the manuscript, however, require further improvement. Major: • Both the extent of the quality gap and its variation between services (service categories) is contingent on the choice of indicators adopted. This implies that to be informative the evaluation should clearly motivate the choice of indicators adopted; in particular, the authors ought to articulate what operational bottlenecks are captured by the process indicators selected, how these relate to health outcomes, and, finally, what should the appropriate policy response be. • There is inconsistency in the scope of quality indicators selected for curative and preventive interventions. While the prior represent process and input indicators (provider), the latter represent service outcomes (patient). It is not clear to me whether these can be effectively combined or compared. • While authors speculate about the relative contribution of process and input indicators to overall performance metrics assessed (page 13, lines 16-23) , they do not, however, evaluate these. It is not clear why this decision was made. Process and input service indicators prompt a different

	policy responses. Understanding how each contributes to the overall performance indicator is thus highly relevant. Each should be reported separately before being aggregated to the overall facility performance metric (applies to Tables 1-A and B). • There is an interesting point in the discussion regarding the qualification of staff providing consultations (page 13, lines 8-30). It seems these data were collected, but not included as one of the service dimensions. Adding this indicator in the manner shown in Nesbitt et al would greatly improve the scope of the effective coverage metrics assessed. • Authors ought to be careful to frame their results and discussion in the context of the sample selected for the analysis. The discussion section needs to consider the implications of the sampling scheme on the findings and generalizability of the estimates it yielded. Finally, there needs to be a clarification of what aspects of preventive and curative CHS are NOT captured by the service indicators selected and to what extent services assessed represent these broader service categories. • The last sentence of the conclusion is highly speculative suggesting that improving effectiveness, as defined by indicators adopted in this study, would not require additional resources. I would urge the authors to be explicit on the interventions that could accomplish that. Minor: • Authors ought to provide further details on the primary health facilities; i.e. how are these facilities staffed, what is the level of training of staff servicing these, number of staff, what these facilities represent (i.e. stand-alone brick building, shack, CHW home), and how important are these as a source of CHS in the study setting (i.e. primary source of care or most children are treated outside of formal sector or through private or NGOs). It would also be very useful if authors commented on how homogenious these health facilities are in the study area. • Authors ought to provide further details on selection criteria of districts included in the study. They indicate these were selected based on "poor performance of certain maternal and child health indicators" (page 5, lines 31-32) but are neither explicit of what "performance" refers to nor clarify what these indicators were. This is important especially as facilities appear to score quite high on the signal functions adopted in this study. • Effective coverage has been conceptualized by a number of related, yet, somewhat different frameworks; each with its nomenclature and target outcome. The authors seem to borrow from a number of the methodologies causing some confusion. I urge that the text is reviewed for consistency of terminology within the adopted framework.

	- For instance, “service utilization” in Tanahashi refers to the relationship between service capacity and output; in Shengelia et al it refers to population take-up on services conditional on need - Authors cite WHO for the definition of effective coverage (page 4, lines 25-29), and then proceed with a definition from Shengelia et al that differs from one adopted in WHO UHC monitoring report. The latter also incorporates the timeliness of service. - On data collection, authors should state whether facility checklists were validated in any way (ie. checks of inventory recorded). If the data were not checked, the implications and potential bias and direction need to be recognized and discussed - The definition for the term Uij in the effective coverage formula (p.7 line 4) is inaccurate; the term is defined differently further in the text (p. 7 line 19-20). Please, correct - For curative CHS authors ought to provide further details on the survey instruments collecting information on illness episodes and health seeking. In particular, the respective survey questions asked of respondents should be listed. - For preventive CHS why partial list of EPI vaccines was assessed? Why not OPV or TT? - On effective coverage for preventive care; authors listed 2 quality indicators: availability of child’s vaccination booklet in the household and timing of vaccination (page 8, lines 37-57); yet, on the following page effective coverage is defined only as a function of vaccination timing (page 9, lines 1-8). Presumably this is assessed in a subset of children with a vaccination booklet. Please, make it explicit. - Results section needs to be re-worked. It describe tables. I would suggest the authors point the reader toward the key findings, align these with take-away messages. - For Figure 2, state the universe (i.e. coverage indicators refer to children that reported an illness (4 wks recall) for curative interventions and age ranges for each vaccine and availability of the vaccination handbook). - The contribution of individual indictors to loss in effective coverage for preventive interventions needs to be demonstrated. It seems that low effective coverage for vaccination is largely due to absence of the vaccination book. Is my reading correct? - Crude coverages for a range of service functions are quite high; this is also an interesting and a positive finding. These are somewhat surprising as the study was conducted in poor-performing districts - Another issue with sample selection that is not noted until discussion (page 14, lines 13-23) is that children without vaccination books are excluded from the analysis. It is not clear whether immunization rates were only assessed among children with vaccination books as well or did this

	indicator include all children and relied on self-report for crude coverage. I would argue that these children should be included otherwise estimates of crude coverage are overstated. I wish the authors the best should they choose to revise the manuscript.

REVIEWER	Drissa Sia Université du Québec en Outaouais Campus de Saint Jérôme, Québec, Canada
REVIEW RETURNED	31-Aug-2017

GENERAL COMMENTS	- Major Compulsory Revisions 1. Methods, page 5, Study Design: this section is more than the Design of the study. I suggest separating it in two section: The Design section and Study participants 2. I am wondering if each health facility (CSPS) had just one provider that interact with patients. Some time you have in the same CSPS a nurse who is the chief of the CSPS and nurse assistant and both provide curative and preventive care to patients. a) So, you should explain what was the situation in these CSPS included in your study. b) You should also explain why you just considered the performance of management of common childhood diseases and severe childhood diseases obtain from only one provider of CSPS, (if they are two providers for example) to qualified the performance of their health facility? Is that correct is the observed provider is nurse assistant because the nurse was for example in training session? 3) Results, page 11, ligne 31, It's not clear in the methods section what you mean by crude and effective coverage for vaccination. 4) Data Analysis: You should describe your data analysis - Minor Essential Revisions 1. Methods, page 5,"Study Design the following section should be part of "Study Setting" "Districts were selected……..91% of these selected facilities were located rurally". 2. Methods, page 5, Study Design, you wrote: "Subsequently, 15 households were randomly selected from all the households". You should inform readers about this random process. 3. Methods, page 6, Data Collections, you wrote "For each U5YO case, the patient provider interaction during the consultation was directly 20 22 21 observed and recorded"; it will be interesting to inform readers, if providers had known that they were observing.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 Major comments

1- **Both the extent of the quality gap and its variation between services (service categories) is contingent on the choice of indicators adopted. This implies that to be informative the evaluation should clearly motivate the choice of indicators adopted; in particular, the authors ought to articulate what operational bottlenecks are captured by**

the process indicators selected, how these relate to health outcomes, and, finally, what should the appropriate policy response be.

Response to the reviewer*:* As suggested by the reviewer 1 we reviewed the discussion section and pointed out this issue **from page 11 line 25 to page 12 line 6**. We agree that effective coverage estimates are heavily influenced and can be easily modulated depending on the indicators selected to measure service quality. Although the process, input, and structural indicators included in our quality score are informed by the work of other authors, they still can be considered selective or biased towards technical elements of the care delivery process. Still, we understand that for health care provision to be effective, evidence-based clinical protocols (such as Integrated Management Childhood Illness (IMCI)) need to be adhered to and can therefore be considered the gold standard against which quality should be measured.

2- **There is inconsistency in the scope of quality indicators selected for curative and preventive interventions. While the prior represent process and input indicators (provider), the latter represent service outcomes (patient). It is not clear to me whether these can be effectively combined or compared.**

Response to the reviewer: We agree on this comment. In the present manuscript we have only kept the quality indicators for curative interventions. We made this decision because combining effective coverage for curative and preventive child health services in the same paper can be overwhelming and difficult to follow for the reader. Hence, we only focus on curative child health services.

3- **While authors speculate about the relative contribution of process and input indicators to overall performance metrics assessed (page 13, lines 16-23) , they do not, however, evaluate these. It is not clear why this decision was made. Process and input service indicators prompt a different policy response. Understanding how each contributes to the overall performance indicator is thus highly relevant. Each should be reported separately before being aggregated to the overall facility performance metric (applies to Tables 1-A and B).**

Response to the reviewer: To measure the overall performance of each indicator we have evaluated both process and inputs indicators even if we did not present the latter in Table1-A and B. We agree with the reviewer that presenting them separately can be relevant for policy responses. Therefore, we changed Table1-A and Table1-B.

4- **There is an interesting point in the discussion regarding the qualification of staff providing consultations (page 13, lines 8-30). It seems these data were collected, but not included as one of the service dimensions. Adding this indicator in the manner shown in Nesbitt et al would greatly improve the scope of the effective coverage metrics assessed.**

Response to the reviewer: We thank the reviewer for her suggestion. We have adjusted the quality score of each health facility by staffing (see Table 2, page 21)**.** We have considered the characteristics of health professionals who performed the consultations and the vignettes. We used healthcare workers' qualification and training to IMCI guidelines.

5- **Authors ought to be careful to frame their results and discussion in the context of the sample selected for the analysis. The discussion section needs to consider the implications of the sampling scheme on the findings and generalizability of the estimates it yielded. Finally, there needs to be a clarification of what aspects of preventive and curative CHS are NOT captured by the service indicators selected and to what extent services assessed represent these broader service categories.**

   Response to the reviewer: As suggested by the reviewer we discussed the implications of the sampling and the generalizability of the findings in the discussion section (from page 12 line 33 to page 13 line 3): While our study focus was on primary level health care facilities in rural areas, study regions and districts were purposely selected, which limits the generalizability of our results. However the large sample available to this study (representing around one third of primary level health care facilities in the country) still provides a sufficiently representative overview on effective coverage in Burkina Faso

6- **The last sentence of the conclusion is highly speculative suggesting that improving effectiveness, as defined by indicators adopted in this study, would not require additional resources. I would urge the authors to be explicit on the interventions that could accomplish that.**

   Response to the reviewer: We agree with the reviewer and we reviewed consequently the conclusion (see page 13 lines 11-21).

Minor

1- **Authors ought to provide further details on the primary health facilities; i.e. how are these facilities staffed, what is the level of training of staff servicing these, number of staff, what these facilities represent (i.e. stand-alone brick building, shack, CHW home), and how important are these as a source of CHS in the study setting (i.e. primary source of care or most children are treated outside of formal sector or through private or NGOs). It would also be very useful if authors commented on how homogenious these health facilities are in the study area.**

   Response to the reviewer: As suggested by the reviewer we have re-written the study setting section. We provided more details on the characteristics of health facilities of the study area (see page 5 lines 3-14).

2- **Authors ought to provide further details on selection criteria of districts included in the study. They indicate these were selected based on "poor performance of certain maternal and child health indicators" (page 5, lines 31-32) but are neither explicit of what "performance" refers to nor clarify what these indicators were. This is important especially as facilities appear to score quite high on the signal functions adopted in this study.**

   Response to the reviewer: As suggested we have clarified this selection in the methods section of the present manuscript (page 5, lines 15-21): Regions and districts have been purposely selected by the government on the basis of poor maternal and child health indicators outcomes: (i) contraceptive prevalence rate: (ii) assisted deliveries; (iii) antenatal consultations (iv) post-natal consultations v) childhood vaccination coverage. This choice was more based on maternal health indicators rather than child health indicators.

3- **Effective coverage has been conceptualized by a number of related, yet, somewhat different frameworks; each with its nomenclature and target outcome. The authors seem to borrow from a number of the methodologies causing some confusion. I urge that the text is reviewed for consistency of terminology within the adopted framework.**

o For instance, "service utilization" in Tanahashi refers to the relationship between service capacity and output; in Shengelia et al it refers to population take-up on services conditional on need
o Authors cite WHO for the definition of effective coverage (page 4, lines 25-29), and then proceed with a definition from Shengelia et al that differs from one adopted in WHO UHC monitoring report. The latter also incorporates the timeliness of service.

Response to the reviewer: As suggested, we reviewed the text according to the definition from Shenglia et al.

4- **On data collection, authors should state whether facility checklists were validated in any way (ie. checks of inventory recorded). If the data were not checked, the implications and potential bias and direction need to be recognized and discussed**

Response to the reviewer: Yes, facility checklists were validated (page lines 17-24): A facility inventory was conducted at each sampled facility assessing the availability of staff, infrastructure, equipment, drugs, supplies, and consumables. Each facility head verbally completed a structured checklist and a research assistant verified availability and functionality of reported items.

5- **The definition for the term Uij in the effective coverage formula (p.7 line 4) is inaccurate; the term is defined differently further in the text (p. 7 line 19-20). Please, correct**

Response to the reviewer: We thank the reviewer for this comment. We took it into account in the present manuscript.

6- **For curative CHS authors ought to provide further details on the survey instruments collecting information on illness episodes and health seeking. In particular, the respective survey questions asked of respondents should be listed.**

Response to the reviewer: For data collection we used Health Results Innovation Trust Fund (HRITF) survey instruments which have been adapted to Burkina context. We added these instruments as additional files to this manuscript.

7- **For preventive CHS why partial list of EPI vaccines was assessed? Why not OPV or TT?**
Response to the reviewer: As we removed from the present manuscript the part regarding preventive child health services, we did not address this question**.**

8- **On effective coverage for preventive care; authors listed 2 quality indicators: availability of child's vaccination booklet in the household and timing of vaccination**

(page 8, lines 37-57); yet, on the following page effective coverage is defined only as a function of vaccination timing (page 9, lines 1-8). Presumably this is assessed in a subset of children with a vaccination booklet. Please, make it explicit.

Response to the reviewer: As we removed from the present manuscript the part regarding preventive child health services, we did not address this question**.**

9- **Results section needs to be re-worked. It describes tables. I would suggest the authors point the reader toward the key findings, align these with take-away messages.**

Response to the reviewer: we thank the reviewer for his suggestion. We have taken it into account in the results section.

10- **For Figure 2, state the universe (i.e. coverage indicators refer to children that reported an illness (4 wks recall) for curative interventions and age ranges for each vaccine and availability of the vaccination handbook).**

Response to the reviewer: As we removed from the present manuscript the part regarding preventive child health services, we did not address this question

11- **The contribution of individual indicators to loss in effective coverage for preventive interventions needs to be demonstrated. It seems that low effective coverage for vaccination is largely due to absence of the vaccination book. Is my reading correct?**

Response to the reviewer: As we removed from the present manuscript the part regarding preventive child health services, we did not address this question

12- **Crude coverages for a range of service functions are quite high; this is also an interesting and a positive finding. These are somewhat surprising as the study was conducted in poor-performing districts.**

Response to the reviewer: As specified above we think that the range of service function are high because regions and districts were selected more based on maternal than children indicators. In 2010, the crude coverage of under-five was 50% (DHS). This percentage has probably increased.

13- **Another issue with sample selection that is not noted until discussion (page 14, lines 13- 23) is that children without vaccination books are excluded from the analysis. It is not clear whether immunization rates were only assessed among children with vaccination books as well or did this indicator include all children and relied on self-report for crude coverage. I would argue that these children should be included otherwise estimates of crude coverage are overstated.**

Response to the reviewer: As we removed from the present manuscript the part regarding preventive child health services, we did not address this question

Reviewer 2 Major comments

1. **Methods, page 5, Study Design: this section is more than the Design of the study. I suggest separating it in two section: The Design section and Study participants**

   Response to the reviewer: We thank the reviewer for his comment. In this study we used different datasets (health facility and household surveys). We think that describing the study's design and participants in the same section will allow the reader to better understand the sampling. In page 5 line 15 we have added the title *study participants* to make it clearer.

2. **I am wondering if each health facility (CSPS) had just one provider that interact with patients. Some time you have in the same CSPS a nurse who is the chief of the CSPS and nurse assistant and both provide curative and preventive care to patients.**
   a) So, you should explain what was the situation in these CSPS included in your study.
   b) You should also explain why you just considered the performance of management of common childhood diseases and severe childhood diseases obtain from only one provider of CSPS, (if they are two providers for example) to qualified the performance of their health facility? Is that correct is the observed provider is nurse assistant because the nurse was for example in training session?

   Response to the reviewer: We thank the reviewer for his comments.

   a) In the current manuscript we described the staffing of health facilities in the study area. In our study the CSPS facilities were composed of Nurse, midwife/assistant and nurse assistant called *Agent Itinérant de Santé*. If the two first providers are qualified to perform under-five year old consultations, the latter is not. But in reality he performs these consultations.

   b) To measure the CSPS facility overall performance we have considered three dimensions of quality:
   - *Management of common childhood diseases*: Trained data collectors observed the U5YO consultations with a checklist based on IMCI guidelines. In the same health facility sometimes these consultations were performed by different health providers.
   - *Management of severe childhood diseases*: For this dimension, the clinical staff present the day of the visit was interviewed to assess their knowledge. We used vignettes on severe pediatric cases such as dehydration, fever, respiratory distress.
   - *General service readiness*: based on the facility inventories including five indicators assessing the availability of electricity, water, sanitation, patient transport, and waiting rooms.

3. **Results, page 11, ligne 31, It's not clear in the methods section what you mean by crude and effective coverage for vaccination.**

Response to the reviewer: As we removed from the present manuscript the part regarding preventive child health services, we did not address this question.

4. **Data Analysis: You should describe your data analysis**
   Response to the reviewer: All analyses were descriptive. In the current manuscript at the section *Measures and analysis* (page 7 line 6), we described our data analysis.

Reviewer 2 Minor Essential Revisions

1. **Methods, page 5,"Study Design the following section should be part of "Study Setting" "Districts were selected……..91% of these selected facilities were located rurally".**

   Response to the reviewer: We thank the reviewer 2 for this comment. As we added study participants in the title, we chose to keep it in this section because this sentence described the participants of the study**.**

2. **Methods, page 5, Study Design, you wrote: "Subsequently, 15 households were randomly selected from all the households". You should inform readers about this random process.**

   Response to the reviewer: We thank the reviewer for this comment that we have taken into account (See page 6 lines 9-11).

3. **Methods, page 6, Data Collections, you wrote "For each U5YO case, the patient provider interaction during the consultation was directly observed and recorded"; it will be interesting to inform readers, if providers had known that they were observing.**

   Response to the reviewer: We thank the reviewer for his comment. Yes, before the direct observation, the provider was informed and we obtained consent from both the provider and the child caretaker. In addition, in the discussion section (page 12 lines 21-24) we shown as one limitation of the study the Hawthorne effect (a common bias of direct observation), which describes higher performance under observation compared to non-observed situations, and may cause overestimation of actual performance.

VERSION 2 – REVIEW

REVIEWER	Andrea Melberg MD/PhD Candidate Centre for International Health University of Bergen Norway E-mail: andrea.melberg@uib.no
REVIEW RETURNED	27-Nov-2017

GENERAL COMMENTS	I thank the authors for the opportunity to review this interesting and very comprehensive study of child health services in Burkina Faso. Even though the article seems to have been through quite extensive

	changes based on the first round of reviews, I still have a number of comments. Major comments 1. I generally think the paper needs to provide a more comprehensive description of the Burkinabè context in the study setting section by for example providing some key information on geography, economy and demography of the country (for example HDI, MMR, U5MR, NNMR, leading causes of deaths in U5YO). 2. More information and background references on the use of vignettes need to be provided. How and by whom were they selected and developed? What kind of standards, protocol or guidelines are they based on? Are they validated and adapted to the Burkinabè context? 3. The measures of both crude and consequently the effective coverage of curative CHS are contingent on the definition of an illness episode, which is not appearing in the manuscript. Currently, the crude coverage seems based on the assumption that all illness episodes reflect a need of health service at the CSPS, and the absence of CSPS treatment thus acts as a proxy for reduced access to care. As a consequence, one might envisage that a child with a simple fever treated at home will result in reduced crude coverage, but that this home treatment does not necessarily reflect a limited access to CHS. 4. I echo the previous comments by reviewer one regarding the motivation of the choice of dimensions and indicators, that I think needs to be further addressed and included the methods section. One suggestion is to clarify how you define quality of care, and how quality is reflected in the three dimensions of care chosen in this article. In addition, the choice and the implication of the threshold levels for the different dimensions of care need to be discussed, as they are pivotal for the CSPS being defined as low or high quality. 5. There exists a vast literature on the effectiveness of the IMCI guidelines and on the reasons for low adherence to these that I suggest to engage with in the discussion of the findings. 6. In the discussion (p11,l32-33) you mention quality differences across CSPS within and across the measured dimension. Based on my limited experience from the Cascades region, there is great variation between rural CSPS in terms of population covered, available infrastructure, and number and formal training of health workers. It would be fruitful, as mentioned by reviewer one, to further elaborate on these variations in the results and discussion sections. 7. The study reports from nearly 1/3 of the primary health centres in Burkina Faso, and the authors claim that the study, due to its large sample, can claim to be “a sufficiently representative overview on effective coverage in Burkina Faso”. As the CSPS surveyed were purposively selected based on poor performance on certin indicators, I find it difficult to be convinced that the findings are representative for all the CSPS of the country or even all rural CSPS. However, I think the findings are still of great importance even though they are neither representative nor generalizable. Minor comments 1. In the abstract you mention “a set of 25 functions”, but these functions do not clearly appear in the methods section. 2. Abstract (p2,l18). According to Figure 2, 5.3% and 44.6% of all children with an illness episode received high or high and intermediate quality respectively, not of the “children who sought care”.

	3. In the introduction, you mention neonatal complications as main reasons of child mortality. However, I cannot see that you address these perinatal health services in your study. If not, this is a major study limitation that needs to be addressed. 4. Methods (p6,l6-12). By selecting 15 households from 494 CSPS you would have reached 7410, not 7694 households. Where are the 184 additional households coming from? 5. While reading the methods section, I am uncertain whether the case samples are collected from all the surveyed CSPS or from a selected subsample. If the latter is the case, this needs to be clearly stated and the selection criteria need to be spelled out. 6. The term "qualified personnel" employed in the study seems to be a result of the MoH definition. Does it exist an international standard for IMCI providers as national and international standards do not always converge ("accoucheuses auxiliaires" are for example defined as skilled birth attendants by the MoH, but not by the WHO). 7. Discussion (p12,l28-30) The two sentences starting with "In addition, to estimate…" needs to reformulation to enhance clarity. 8. Conclusions of abstract and main text: Two out of three service criteria selected in this study (MCCD and MSCD) are heavily based on the adherence to guidelines and protocols. Your conclusion that "poor adherence clinical treatment guidelines seemed to be the main contributors to the gap between crude and effective coverage" appears to me as a consequence of your choice of indicators. The statement thus becomes a tautology. 9. Conclusion: I recommend reviewing the choice of the wording "enforcement".

REVIEWER	Drissa sia Université du Québec en Outaouais (UQO), Canada
REVIEW RETURNED	29-Nov-2017

GENERAL COMMENTS	The authors have correctly addressed my comments

REVIEWER	Thomas Druetz Tulane University, USA
REVIEW RETURNED	01-Dec-2017

GENERAL COMMENTS	This manuscript assesses quality of healthcare services provided in rural health facilities in Burkina Faso. Based on this assessment, the authors adjusted the crude coverage for curative health services. The study is highly relevant, its methods are sound and results are very interesting. We believe the manuscript would be improved if the following comments are addressed. • The use of the terminology "effective coverage" is unfortunate. The authors acknowledge in their conclusion (p. 12, l. 28-32) that they "used only indicators of content of care to assess a potential health gain" that "may not directly translate into health gain". They further state that their indicators "did not capture patients' adherence to treatment or individual health outcomes". In this context, it is debatable to refer to effective coverage in this study. It seems to be a coverage indicator adjusted for quality of care. Similarly, the expressions "effectiveness of care" and "effectiveness of services" seem really inappropriate here, since there is no measure whatsoever on the effects of services provided

	at health facilities. • External validity: P. 5, L. 19, the manuscript states that districts were selected “on basis of low performance”. Please justify why. Also, this is an important limit to the generalizability of the results. On page 12, the authors acknowledge this limit but then argue (p. 13, lines 1-3) that the sample of health facilities was large, which compensates for that limitation. It does not. Take for example a study that enrolls 50% of the total population; if the sampling is made among males only, results can hardly be generalized to females. • The use of different thresholds for defining the categories of quality is unclear (Table 2). Why a HF's performance is high if it has a score ≥ 7 for MCCD, ≥ 8 for MSCD, and ≥ 4 for service readiness? How were the different thresholds defined? Minor comments • P4, L14: consider changing “as long as” by “especially if”. The fact that quality remains poor does not automatically imply that health outcomes will not improve; and the fact that quality improves does not automatically imply that health outcomes will improve. • P4, L18-25: any example from Burkina Faso? It would be useful to present results from previous studies that assessed the crude coverage of curative care for children <5 in Burkina Faso. See for example: o Druetz T et al. (2015). Community case management of malaria: results from a three-year panel study of treatment-seeking behavior in the districts of Kaya and Zorgho, Burkina Faso, Malaria Journal, 14(71). o Druetz T et al. (2015). Abolishing fees at health centers in the context of community case management of malaria: What effects on treatment-seeking practices for febrile children in rural Burkina Faso?, PLoS ONE, 10(10): e0141306. o Several studies from Nouna research center. • P4, L26: I am confident that the study objective was more than simply “contribute to the effective coverage literature”. Please rephrase and state that this type of assessment of quality-adjusted coverage for curative services has never been conducted in Burkina Faso, or any other objective that the authors pursued for that matter. • P6, L10: what was the sample frame (and how was it obtained) used to randomly select households? Why 15 households were selected per village? Please justify. • P6, L25: please provide in annex the IMCI algorithm used in Burkina Faso for standard case management. • P7, L1-5: structured interviews were conducted to pursue what objective? Where are the results? • P8, L11: how was the adjustment performed? We do not see any adjusted results, and the authors state the analyses are only descriptive. Please clarify. • P8, L22: please justify why effective coverage was not computed for all 3 categories of performance. • P9, L7: “in” twice. • Discussion: it would be interesting to compare the adherence to the algorithm observed in this study at health facilities to the adherence that CHW show when performing integrated-community case management. • Table 1A: “Due to generally low performance of this indicator if measured against this standard, we considered this process to be performed when at least two danger signs were reviewed.” One can argue that performance indicators should not be changed in order to get higher scores. If this remains, it is important

	to mention it clearly in the discussion and limitations sections. It is an interesting result that a very low percentage of providers ask for the four danger signs as per requested by the IMCI guidelines. • Tables 1a and 1B: please remind the reader that the performance was directly observed for the former, and based on scenario vignettes for the second.

VERSION 2 – AUTHOR RESPONSE

Dear Editors and reviewers

Please find below our response:

Editorial Requests:

- Please revise your title to indicate the research question, study design, and setting. This is the preferred format of the journal.

Response to Editors: We changed the title to better align with the journal format.

- The strengths and limitations section (page 3) needs improving. The first two bullet points do not relate to the methods or design of the study.

As a reminder, this section should contain up to five short bullet points, no longer than one sentence each, that relate specifically to the methods/ design of the study reported (see: http://bmjopen.bmj.com/site/about/guidelines.xhtml#articletypes).

Response to Editors: We improved this section according to the comments (Page3; Lignes1-11).

Reviewers' Comments to Author:

Reviewer: 1

Reviewer Name: Andrea Melberg

Institution and Country: MD/PhD Candidate, Centre for International Health, University of Bergen, Norway

Competing Interests: None declared

I thank the authors for the opportunity to review this interesting and very comprehensive study of child health services in Burkina Faso. Even though the article seems to have been through quite extensive changes based on the first round of reviews, I still have a number of comments.

Response to reviewer: We thank reviewer 1 for her valuable comments. We outline below in detail how we decided to address the points raised by the reviewer.

Major comments

1. I generally think the paper needs to provide a more comprehensive description of the Burkinabè context in the study setting section by for example providing some key information on geography, economy and demography of the country (for example HDI, MMR, U5MR, NNMR, leading causes of deaths in U5YO).

Response to reviewer: We revised the study setting section by listing additional key features on the country context relevant to our study's background (Page 5, lines3-7).

2. More information and background references on the use of vignettes need to be provided. How and by whom were they selected and developed? What kind of standards, protocol or guidelines are they based on? Are they validated and adapted to the Burkinabè context?

Response to reviewer: The vignettes used in our study are based on IMCI guidelines and vignettes developed by Shivam Gupta and David Peters from John Hopkins University and included in the impact evaluation toolkit produced by the Health, Nutrition and Population Hub (HDNHE) in collaboration with regional RBF teams and the Development Economics Research Group (DECRG). We adjusted in the text the following references.

(http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTHEALTHNUTRITIONANDPOPULATION/EXTHSD/EXTIMPEVALTK/0,,contentMDK:23262154~pagePK:64168427~piPK:64168435~theSitePK:8811876,00.html). We adapted them to the Burkina Faso context taking into account the national standard treatment guidelines. The vignettes are not validated, but have been included in RBF impact country evaluations.

3. The measures of both crude and consequently the effective coverage of curative CHS are contingent on the definition of an illness episode, which is not appearing in the manuscript. Currently, the crude coverage seems based on the assumption that all illness episodes reflect a need of health service at the CSPS, and the absence of CSPS treatment thus acts as a proxy for reduced access to care. As a consequence, one might envisage that a child with a simple fever treated at home will result in reduced crude coverage, but that this home treatment does not necessarily reflect a limited access to CHS.

Response to reviewer: The reviewer points towards the difference between perceived need (i.e. perception of health risk and belief that using health service will make a difference) and true need (i.e. actual illness requiring health service use regardless of perceived need).

Our approach agrees with the reviewer that perceived need doesn't necessarily equal true need, as perceived need might also include individuals without true need. Our definition of utilization conditional on true need therefore followed the underpinnings by Shengelia et al 2005 to the extent possible given the data available to us: our best approximation of true need is the report on an illness episode. Whether service utilization occurred or not, is a function of perceived need among those with assumed true need. Hence, the best approximation of U | N in our study is service utilization given true need (defined as reported illness episode).

We clarified our definition and assumptions regarding N and U | N more explicitly (P7; L16-26).

4. I echo the previous comments by reviewer one regarding the motivation of the choice of dimensions and indicators, that I think needs to be further addressed and included the methods section. One suggestion is to clarify how you define quality of care, and how quality is reflected in the three dimensions of care chosen in this article. In addition, the choice and the implication of the threshold levels for the different dimensions of care need to be discussed, as they are pivotal for the CSPS being defined as low or high quality.

Response to reviewer: The choice of quality dimensions and indicators used in this study are based on Donabedian's quality of care framework and the quality of child health care measures suggested by Gouws et al. According to Donabedian, quality of care can be assessed along structural elements (e.g. staff, infrastructure, equipment…), process elements (i.e. interaction between patients and providers) and outcome elements (e.g. patient satisfaction). According to Gouws et al. quality of child health care at first line health facilities can be defined along four indices: 1) Integrated child assessment (based on IMCI guidelines) 2) Facility readiness to deliver IMCI, 3) Capacity to manage severe illness using vignettes, 4) Capacity to manage severe illness given availability of essential drugs. These indices are sufficiently validated.

Our quality dimension MCCD is based on indices 1 and 2; the MSCD dimension is based on indices 3 and 4; the general facility readiness dimension represents structural elements relevant to essential facility infrastructure and base on the Donabedian framework.

We agree that quality of care as a multidimensional concept is challenging to assess and thus deserves further attention. We therefore revised the relevant sections in the methods chapter (P7, L27-33 and P8, L1-16).

We further included a more detailed discussion on the potential limitations of our definition and selection of quality dimensions and category thresholds (P12; L25-31 and P14, L1-10).

5. There exists a vast literature on the effectiveness of the IMCI guidelines and on the reasons for low adherence to these that I suggest to engage with in the discussion of the findings.

Response to reviewer: In the discussion section (P13, L10-23) we have tried to outline the literature on IMCI guideline adherence in low income settings a bit more prior to pointing at the more specific reasons we feel responsible in the Burkina context (lack of IMCI training for the AIS in spite their use in service provision, lack of essential equipment).

6. In the discussion (p11,l32-33) you mention quality differences across CSPS within and across the measured dimension. Based on my limited experience from the Cascades region, there is great variation between rural CSPS in terms of population covered, available infrastructure, and number and formal training of health workers. It would be fruitful, as mentioned by reviewer one, to further elaborate on these variations in the results and discussion sections.

Response to reviewer: We clarified this sentence more (P12; L32-33). In the revised manuscript we pointed out the variation between the three quality dimensions (MCCD, MSCD, general facility readiness) as showed in Figure1. Facilities tend to have higher quality score for facility readiness than MCCD or MSCD. While our study revealed that most of rural facilities had access to basic infrastructures, some structural differences may still remain. In addition, the variation between the three quality dimensions abovementioned may due to differences on facilities characteristics (type of population covered, location, etc.) not picked up by our survey.

7. The study reports from nearly 1/3 of the primary health centres in Burkina Faso, and the authors claim that the study, due to its large sample, can claim to be "a sufficiently representative overview on effective coverage in Burkina Faso". As the CSPS surveyed were purposively selected based on poor performance on certain indicators, I find it difficult to be convinced that the findings are representative for all the CSPS of the country or even all rural CSPS. However, I think the findings are still of great importance even though they are neither representative nor generalizable.

Response to reviewer: We agree with the reviewer comment. Hence we revised the statement on generalizability in the discussion section (P14, L19-24).

Minor comments

1. In the abstract you mention "a set of 25 functions", but these functions do not clearly appear in the methods section.

Response to reviewer: We agree that the abstract is somewhat misleading, as our focus is not so much the 25 functions, but rather the approach to determining a quality score applicable to estimate effective coverage. The 25 indicators or functions are still sufficiently visible in Tables 1A-C without deserving further emphasis in the abstract. We therefore removed "25 functions" from the abstract (P2, L8) and talk simply of a "set of indicators".

2. Abstract (p2,l18). According to Figure 2, 5.3% and 44.6% of all children with an illness episode received high or high and intermediate quality respectively, not of the “children who sought care”.

Response to reviewer: This is an oversight on our part and we appreciate this having been pointed out. We have corrected this error in the abstract.

3. In the introduction, you mention neonatal complications as main reasons of child mortality. However, I cannot see that you address these perinatal health services in your study. If not, this is a major study limitation that needs to be addressed.

Response to reviewer: While our third vignette was a case of lethargic 1 month old (P8, L1 and in Table1-B) we did not address perinatal conditions since our focus was on U5YO and infants. Hence, we removed in the introduction the wording referring to neonatal conditions (P4, L5-6) and we pointed out this as limitation (P14, L18-19).

4. Methods (p6, l6-12). By selecting 15 households from 494 CSPS you would have reached 7410, not 7694 households. Where are the 184 additional households coming from?

Response to reviewer: We thank the reviewer for this comment. Indeed, we made a mistake in representing the real number of households. The correct number of households surveyed in this study is 7347. We wrongly included households surveyed in CSPS catchment areas that we specifically excluded from the facility sample (see criteria P5; L26-30). We don’t have the 7410 households because 4 villages (60 households) related to 4 CSPS included in the facility sample were not surveyed and in 3 villages only 14 household instead of 15 were surveyed. The number of U5YO is the right figure.

5. While reading the methods section, I am uncertain whether the case samples are collected from all the surveyed CSPS or from a selected subsample. If the latter is the case, this needs to be clearly stated and the selection criteria need to be spelled out.

Response to reviewer: For this study we have excluded 19 facilities that did not meet our criteria as defined in the methods section (P5; L30-32): Recently opened facilities (less than six months old) or other forms of primary care services (e.g. at high schools, colleges, garrisons or, prisons). However, it seems that the definition of this subsample was not sufficiently clear. We therefore rephrased slightly to make this point more obvious (P5, L28-33).

6. The term “qualified personnel” employed in the study seems to be a result of the MoH definition. Does it exist an international standard for IMCI providers as

national and international standards do not always converge (“accoucheuses auxiliaires” are for example defined as skilled birth attendants by the MoH, but not by the WHO).

Response to reviewer: As correctly pointed out, national and international standards do not always align. We therefore purposefully relied on the national (MoH) standards in defining personnel qualification. For instance, the “Agent Itinérant de santé” (AIS) represent a professional cadre rather specific to the Burkina context. The MoH does not consider AIS to perform curative consultation and thus does not include this group in IMCI training programs. The reality at facility levels, as pointed out by our findings, seems to indicate that AIS provide curative services independently albeit their lack of IMCI training.

7. Discussion (p12,l28-30) The two sentences starting with “In addition, to estimate…” needs to reformulation to enhance clarity.

Response to reviewer: We took this comment into account and rephrased the sentence for better clarity (P12; L32-34).

8. Conclusions of abstract and main text: Two out of three service criteria selected in this study (MCCD and MSCD) are heavily based on the adherence to guidelines and protocols. Your conclusion that “poor adherence clinical treatment guidelines seemed to be the main contributors to the gap between crude and effective coverage” appears to me as a consequence of your choice of indicators. The statement thus becomes a tautology.

Response to reviewer: We agree with the reviewer that since our quality dimensions are mainly based on the adherence to guidelines, lack of guideline adherence might be a trivial explanation. However, adherence to evidence-based clinical protocols and guidelines is a central element of clinical care provision, as it guarantees a systematic way to patient care and ensures that no unnecessary or wrong treatments are provided. For this reason, quality of care assessment on observation of guideline adherence (as e.g. done by content of care approaches to quality), are considered highly reliable. If a large proportion of providers disregards guidelines they are supposed to be trained in and that serve as clinical algorithms for quality patient care, this is concerning.

We agree that the next question should be why this is the case and often the answer is poor provider training (awareness of protocols) and inadequate resources at the facility to allow protocols to be fully adhered to. Both of these input elements, knowledge and facility infrastructure, have been assessed and included to the extent relevant in our score computation. We think that we have outlined these links between performance based on guidelines and availability of relevant input items sufficiently already in the results section. For instance, for almost all input elements assessed, availability of these inputs was largely given, still related protocol adherence lacked. To this regard, we think our statement on identifying poor guideline adherence may be the main issue in our setting, has quite some substance.

9. Conclusion: I recommend reviewing the choice of the wording “enforcement”.

Response to reviewer: we replaced it by “implementation”

Reviewer: 2

Reviewer Name: Drissa sia

Institution and Country: Université du Québec en Outaouais (UQO), Canada

Competing Interests: None declared

The authors have correctly addressed my comments

Response to reviewer: We are pleased that our responses addressed the reviewer’s comments sufficiently.

Reviewer: 3

Reviewer Name: Thomas Druetz

Institution and Country: Tulane University, USA

Competing Interests: None declared

This manuscript assesses quality of healthcare services provided in rural health facilities in Burkina Faso. Based on this assessment, the authors adjusted the crude coverage for curative health services. The study is highly relevant, its methods are sound and results are very interesting.

Response to reviewer: We thank reviewer 3 for his valuable comments. We outline below in detail how we decided to address the points raised by the reviewer.

We believe the manuscript would be improved if the following comments are addressed.

• The use of the terminology “effective coverage” is unfortunate. The authors acknowledge in their conclusion (p. 12, l. 28-32) that they “used only indicators of content of care to assess a potential health gain” that “may not directly translate into health gain”. They further state that their indicators “did not capture patients’ adherence to treatment or individual health outcomes”. In this context, it is debatable to refer to effective coverage in this study. It seems to be a coverage indicator adjusted for quality of care. Similarly, the expressions “effectiveness of care” and “effectiveness of services” seem really inappropriate here, since there is no measure whatsoever on the effects of services provided at health facilities.

Response to reviewer: In agreement with the reviewer, effective coverage is a coverage measure that adjusts contact coverage (service use by those in need of a service) by the expected potential of health gain that is received when using a given service. This adjustment represents the effectiveness or quality component added. Tanahashi first introduced this definition in 1978. Since then, there has been a scientific debate on how to best define this effectiveness or quality component to best represent expected health gain (see for example Shengelia et al. 2005, Ng et al 2013). As quality or effectiveness represent constructs that remain challenging to define and measure. Recent literature on effective coverage suggests a number of approaches (e.g. content of care, biomarkers, cohort registration, exposure matching) to approximate the expected health gain (Ng et al 2013,). Each of these approaches has its strength and limitations. In approximating the expected health gain provided by U5YO health services in our study, content of care assessment based on standards determining the readiness of facilities to provide such services and based on the competence of individual providers to comply with guidelines determining optimal clinical performance, seemed to be the best possible approach. A main caveat of quality or effectiveness assessments based on content of care is that these measures do not account for additional aspects related to health gain that are beyond the content of care (which we outlined in our study as a potential limitation).

As we agree with the reviewer’s criticism regarding the definition of effectiveness or quality in general, we still feel confident that the definitions used in our study align to the extent possible with recent and current literature on the specific topic effective coverage,

• External validity: P. 5, L. 19, the manuscript states that districts were selected “on basis of low performance”. Please justify why. Also, this is an important limit to the generalizability of the results. On page 12, the authors acknowledge this limit but then argue (p. 13, lines 1-3) that the sample of health facilities was large, which compensates for that limitation. It does not. Take for example a study that enrolls 50% of the total population; if the sampling is made among males only, results can hardly be generalized to females.

Response to reviewer: For the implementation of Performance based financing (PBF) program in Burkina Faso, the Ministry of Health prioritized districts based on their low performance on maternal and child indicators. Regarding the statement on generalizability, we agree with the reviewers’ comment. Nevertheless we think that the scale of data available to this study (about 500 health facilities and about 7,000 households across six regions), our findings still quite robust and meaningful. Hence we revised the sentence on the generalizability in the discussion section (P14, L19-24).

• The use of different thresholds for defining the categories of quality is unclear (Table 2). Why a HF's performance is high if it has a score ≥ 7 for MCCD, ≥ 8 for MSCD, and ≥ 4 for service readiness? How were the different thresholds defined?

Response to reviewer: These thresholds are not based on any empiric approach and in determining them, we relied on the work of others authors who used arbitrary thresholds to categorized health facilities to assess the effective coverage of maternal health services. For instance Kruk et al 2017 used ≥80% and <50% as thresholds of high and low quality respectively and Nesbitt et al 2013 used ≥60% and <50%. For our study we decided to average the thresholds of these studies and we got: ≥70% and <50% as thresholds of high and low quality respectively.

We elaborated on the implication of these rather arbitrary thresholds in respect to resulting limitations in the revised manuscript (P14, L13-17).

Minor comments

• P4, L14: consider changing "as long as" by "especially if". The fact that quality remains poor does not automatically imply that health outcomes will not improve; and the fact that quality improves does not automatically imply that health outcomes will improve.

Response to reviewer: we agree with this comment. Therefore we have taken it into account in the revised manuscript.

• P4, L18-25: any example from Burkina Faso? It would be useful to present results from previous studies that assessed the crude coverage of curative care for children <5 in Burkina Faso. See for example:

o Druetz T et al. (2015). Community case management of malaria: results from a three-year panel study of treatment-seeking behavior in the districts of Kaya and Zorgho, Burkina Faso, Malaria Journal, 14(71).

o Druetz T et al. (2015). Abolishing fees at health centers in the context of community case management of malaria: What effects on treatment-seeking practices for febrile children in rural Burkina Faso? PLoS ONE, 10(10): e0141306.

o Several studies from Nouna research center.

Response to reviewer: We agree with the reviewer that it is useful to show findings from others studies on crude coverage of U5YO curative services in Burkina Faso. Hence, we have rephrased the last paragraph of the introduction section (P4, L25-27) and referenced the interesting papers shared by the reviewer P4, L26:

• I am confident that the study objective was more than simply "contribute to the effective coverage literature". Please rephrase and state that this type of assessment of quality-adjusted coverage for curative services has never been conducted in Burkina Faso, or any other objective that the authors pursued for that matter.

Response to reviewer: We agree with the reviewer and we rephrased the sentence accordingly (P4, L25-27).

• P6, L10: what was the sample frame (and how was it obtained) used to randomly select households? Why 15 households were selected per village? Please justify.

Response to reviewer: Households within a given catchment area were identified using a two-stage sampling technique. First, one village was randomly selected from all villages located within a given catchment area. Second, in each selected village a household qualified to be included in the sampling frame if the household had at least one pregnant woman or a woman who gave birth within the previous two years. The reason for sampling based on residential pregnant women was that the overall evaluation study also focused on maternal care aspects. This was assumed to still yield a sufficiently representative approach to the U5YO sample, although sampling was restricted to these households defined by recent pregnancies. These household were then listed to constitute the sampling frame, from which 15 households were randomly selected. The reason to randomly select at least 15 households was justified by an ex ante power calculation to detect a minimum acceptable effect size (overall study this data is taken from is on measuring the impact of a results-based financing program on primary care service delivery using a controlled clustered pre-test-post-test design..

• P6, L25: please provide in annex the IMCI algorithm used in Burkina Faso for standard case management.

Response to reviewer: We provided the IMCI algorithm used in Burkina Faso in the annex.

• P7, L1-5: structured interviews were conducted to pursue what objective? Where are the results?

Response to reviewer: Caregivers were interviewed on several aspects related to economic, demographic and health seeking behaviors. Questionnaires and interview questions were primarily developed for the purpose of the overall impact assessment. For this study we only used a minimum sub-set of variables relevant to this study's purpose to define child illness episode (need) and resultant care-seeking behavior (utilization).

• P8, L11: how was the adjustment performed? We do not see any adjusted results, and the authors state the analyses are only descriptive. Please clarify.

Response to reviewer: We wrongly used the wording "adjust" in the sentence. The sentence the reviewer mentioned refers to the categorization of facilities combining their performance scores and their health professionals' characteristics (as shown in table 2). Therefore we rephrased this sentence to clarify better (P8, L30-34).

• P8, L22: please justify why effective coverage was not computed for all 3 categories of performance.

Response to reviewer: To derive a score sufficiently comprehensive in representing a measure of quality or effectiveness of the U5YO health services, we identified these three dimensions (clinical performance and preparedness related to general child care, clinical performance and preparedness related to management of severe childhood illness, general service readiness) as aspects of quality service provision that if optimal is likely to translate into health gain once a service is used by a sick child. Disaggregating these components would reduce this comprehensiveness we were aiming for. Reviewing the literature on U5YO service quality did not yield any further dimensions we could have easily framed with our content of care approach. While we think the performance of facilities within each dimension is useful information – and thus presented in our results section – we don't see an additional benefit in estimating effective coverage for each dimension as this would be somewhat contrary to the concept of effective coverage.

• P9, L7: "in" twice.

Response to reviewer: We have removed the redundant "in".

• Discussion: it would be interesting to compare the adherence to the algorithm observed in this study at health facilities to the adherence that CHW show when performing integrated-community case management.

Response to reviewer: We agree with the reviewer that further investigation of the implementation and adherence to integrated childhood care is needed and relevant, especially in respect to comparisons between facility- and community-based providers. We feel that this is somewhat beyond the focus of this current paper and could probably a publication on its own. While the larger impact study also conducted surveys with CHW, these surveys did not include sufficient data on protocol adherence.

• Table 1A: "Due to generally low performance of this indicator if measured against this standard, we considered this process to be performed when at least two danger signs were reviewed." One can argue that performance indicators should not be changed in order to get higher scores. If this remains, it is important to mention it clearly in the discussion and limitations sections. It is an interesting result that a very low percentage of providers ask for the four danger signs as per requested by the IMCI guidelines.

Response to reviewer: For some indicators performance thresholds had to be lowered to make the overall score work better, which overestimates actual performance. We discussed this limitation in the revised manuscript (P14, L9-13).

• Tables 1a and 1B: please remind the reader that the performance was directly observed for the former, and based on scenario vignettes for the second.

Response to reviewer: We took it into account in the revised manuscript.

VERSION 3 – REVIEW

REVIEWER	Thomas Druetz Tulane University, USA
REVIEW RETURNED	06-Jan-2018

GENERAL COMMENTS	The authors adequately answered to our previous comments, and we found the manuscript very interesting and suitable for publication.

REVIEWER	Andrea Melberg Centre for International Health University of Bergen Norway
REVIEW RETURNED	12-Jan-2018

GENERAL COMMENTS	I want to take the opportunity again to thank the authors for the opportunity to review this very impressive work on crude and effective coverage of sick-child care in primary health centres in Burkina Faso. The study is very relevant and the findings play into the current global discussions the importance of quality of care in the UHC and SDG era. The authors have addressed my comments from the previous round in a very satisfactory manner, and I only have some very minor comments left to the paper: Page5, line 13 and elsewhere: You translate AIS into nurse assistant. As the direct word by word translation of AIS would be something like ambulant health worker, I do recommend you to use another translation of AIS that has been used in the medical literature: Outreach health worker. This term also better describes the intended tasks of AIS, namely vaccinations and other preventive health services in communities. Page 5, line 18: I suggest the authors specify that the subsidization program implies free services for all U5YO. Page 7, line 16-27. Even though this section has been clarified, I still have some concerns about your definition of true need as a reported illness episode. As Shengilla et al mention, intervention-specific coverage studies as the one you have carried out is more complicated than normative ones (ANC, Vaccination, child birth etc), and that true need is defined according to the fact that ”individuals need a health intervention if their expected health gain from receiving it is greater than zero”. I am however unsure whether every U5YO reporting with an illness episode in this study (fewer, diarrhoea etc) would have an expected health gain of more than zero as not all fevers, diarrhoeas or coughs requires medical interventions. If I, for example reported that my 2 year-old had had diarrhoea, but that I have not taken him to the health centre or given him any medical interventions, I would be classified as a non-user, leading to a lower crude and effective coverage for child health services in Norway. I therefore suspect that your definition of illness episode could result in an underestimated crude and hence effective coverage of CHS as reported illness does not always equate with true need. Page 15, line 5-6. You have added a very interesting point in the conclusion regarding the influence of supplies on non-adherence to IMCI-guidelines. These findings are in line with a recently published study by Leslie et al. in Plos Medicine, which shows very limited correlation between structural aspects of care and the process of providing evidence-based maternal and child health care in 8 low-income countries:

	http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002464#pmed.1002464.s003 However, when reading the results section, I can't find reference to the pattern referred to in the conclusion, something I suggest that you do.

VERSION 3 – AUTHOR RESPONSE

Editorial Request:
Please proofread the paper one more time. Page 3: please amend "We conducted this study in around five hundred of primary level health facilities" to "We conducted this study in around five hundred primary level health facilities"

Response to editor: We proofread the paper and we amended the sentence as the editor suggested on page3.

Reviewer: 1

I want to take the opportunity again to thank the authors for the opportunity to review this very impressive work on crude and effective coverage of sick-child care in primary health centres in Burkina Faso. The study is very relevant and the findings play into the current global discussions the importance of quality of care in the UHC and SDG era. The authors have addressed my comments from the previous round in a very satisfactory manner, and I only have some very minor comments left to the paper:

Response to reviewer: We thank reviewer 1 for her comments. We outline below in detail how we decided to address the points raised by the reviewer.

Page5, line 13 and elsewhere: You translate AIS into nurse assistant. As the direct word by word translation of AIS would be something like ambulant health worker, I do recommend you to use another translation of AIS that has been used in the medical literature: Outreach health worker. This term also better describes the intended tasks of AIS, namely vaccinations and other preventive health services in communities.

Response to reviewer: We agree with the suggested translation for this French professional title and concur that this translation captures and reflects the original French meaning more closely. We replaced the translation accordingly on Page5, line 13, where we first introduce the term.

Page 5, line 18: I suggest the authors specify that the subsidization program implies free services for all U5YO.

Response to reviewer: We made the description of the subsidization program more clear by stating that it includes free services for all U5YO, as suggested by the reviewer (Page 5, line 19)

Page 7, line 16-27. Even though this section has been clarified, I still have some concerns about your definition of true need as a reported illness episode. As Shengilla et al mention, intervention-specific coverage studies as the one you have carried out is more complicated than normative ones (ANC, Vaccination, child birth etc), and that true need is defined according to the fact that "individuals need a health intervention if their expected health gain from receiving it is greater than zero". I am however unsure whether every U5YO reporting with an illness episode in this study (fewer, diarrhoea etc) would have an expected health gain of more than zero as not all fevers, diarrhoeas or coughs requires medical interventions. If I, for example reported that my 2 year-old had had diarrhoea, but that I have not taken him to the health centre or given him any medical interventions, I would be classified as a non-user, leading to a lower crude and effective coverage for child health services in Norway. I therefore suspect that your definition of illness episode could result in an underestimated crude and hence effective coverage of CHS as reported illness does not always equate with true need.

Response to reviewer: We understand and appreciate the concern raised by the reviewer. While we believe that our definition of true need by report of an illness episode may underestimate crude and effective coverage is the most appropriate given study context and data available to us, we would like to stress the potential risk of underestimation that could have been incurred with this definition a bit more in the limitation section. (Page14 lines 27-30).

Page 15, line 5-6. You have added a very interesting point in the conclusion regarding the influence of supplies on non-adherence to IMCI-guidelines. These findings are in line with a recently published study by Leslie et al. in Plos Medicine, which shows very limited correlation between structural aspects of care and the process of providing evidence-based maternal and child health care in 8 low-income countries:
http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002464#pmed.1002464.s003
However, when reading the results section, I can't find reference to the pattern referred to in the conclusion, something I suggest that you do.

Response to reviewer: Thank you for pointing this out. We now pointed at this pattern in the results section (page11, lines 5-8) and referred to it in the discussion in respect to the interesting study you suggested (page13, lines 8-12). We now feel that these edits make the point raised in the conclusion section more traceable.

Reviewer: 3
The authors adequately answered to our previous comments, and we found the manuscript very interesting and suitable for publication.

Response to reviewer: We are pleased that our responses addressed the reviewer's comments sufficiently.

VERSION 4 – REVIEW

REVIEWER	Andrea Melberg Centre for International Health, University of Bergen, Norway
REVIEW RETURNED	30-Jan-2018

GENERAL COMMENTS	The authors have addressed my comments in a satisfactory manner.