PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Integration of HIV and maternal health care in a high HIV-prevalence setting: Analysis of client flow data over time in Swaziland
AUTHORS	Birdthistle, Isolde; Mayhew, Susannah; Kikuvi, Joshua; Zhou, Weiwei; Church, Kathryn; Warren, Charlotte; Nkambule, Rejoice; Fenty, Justin

VERSION 1 - REVIEW

REVIEWER	Stephanie Topp Centre for Infectious Disease Research in Zambia
REVIEW RETURNED	04-Oct-2013

GENERAL COMMENTS	The authors have clearly put considerable effort into this manuscript and I applaud their hard work. However, I think the conclusions drawn from the data are frequently overstated, and the manuscript presented requires considerable reworking. Compulsory Revisions Introduction: • Last sentence of para 2, pg 6: the authors refer to 'process evaluations' – but only reference one paper. Is there more than one? • Para 3, pg 6: the authors present findings from a systematic review which concludes that integration of services is 'feasible'. What this does not account for is that the studies reviewed would have shown integration was feasible under certain circumstances. Such findings are never an absolute. When using such studies as background, it is important to qualify the use of such evidence with a demonstration of your understanding that programmatic 'success' is always context specific. • Para 3, pg 6: the second half of this para present the evidence with a little more nuance and qualification; remember, integration can yield benefit, but it is not an automatic outcome of any integration intervention. Previous experience implementing integrated programs elsewhere is not a guarantee. Integration requires a series of complex actions to take place and if any one goes wrong, then it may not have the intended outcome. • General intro: the authors don't define what integration means in the context of your program. Integration can mean different things to different people; for the sake of clarity and precision the authors need to outline what this term means in your setting. Methods: • Pg 7, line 52: how did the authors determine that the control facilities had no integrated HIV-PNC services? This should be clarified in relation to the above suggestion to define what integration means in this setting. • Pg 8, line 5: Integra project was designed to strengthen and maintain the provision of integrated HIV and PNC services – but the authors don't specify what the anticipated benefits of integrated

	(versus non-integrated) service delivery would be. This is important to understanding the study's own objectives. • Pg 10, The very different nature of the control and intervention clinics calls into the question the validity of this comparison. Since the main measure was a facility-specific process (i.e. integrated services) I don't really understand the concern about contamination? o Moreover – it would appear that there were other forms of 'integration' interventions even in the control facilities o The authors need to more clearly outline the contextual / service differences in these facilities – e.g. what service units did each one have; how many professional staff; how many lay staff; what was the catchment population and catchment area; how well supplied were they…all these things would be enormously important to understanding whether and how well services become integrated. • pg 15, lines 33/36: use of the term 'disappeared' is disingenuous. It would seem that healthcare providers simply ceased to deliver integrated services. The outstanding question would appear to be, why? (possible reasons: they felt it was outside their line of duty; they felt it was too much work; they experienced resistance from clients; they didn't have appropriate supplies; they forgot / they were called away or trained to do it differently /new staff transferred in and handed received the orientation). • Simliar questions apply to the findings under sub-heading receipt of HIV counselling. Discussion: • The finding that there was no evidence of increased integration in the intervention clinics appears valid – but I find little additional value to the comparison with control clinics • The absence of any qualitative (or anecdotal programmatic) data with which to meaningfully interpret the quantitative findings is a significant limitation. This means that the authors rely on conjecture to understand the relative lack of integration of service delivery and lack of meaningful patterns in integration across program vs. control clinics. • Pg 18, line 5/6: the authors suggest that 'HIV counselling…has an important role to play in scaling up integration'— in doing so they seem to suggest that integration is an inherent good. This is manifestly not the case. Integration is a means to an end – whether that be more efficient or less stigmatizing health care. Much more care needs to be taken by the authors in considering the implications of their results. I would like to understand if they have prima facae evidence of integration providing programmatic or clinical benefit in the Swazi context – if such evidence does exist, then this must be much more clearly presented, either as part of a literature review or in the results. • Pg 18, lines 37-48: again the authors seem to suggest that the effectiveness of the Integra initiative is a given – and that it is the clinics or context in which they are working that needs to change to demonstrate this effectiveness. The results would suggest that the Integra initiative failed to meaningfully address a range of programmatic, operational and broader contextual factors relating to the provision of integrated MCH services. THIS is a useful and meaningful (even if less palatable) finding in the context of many, similar programs, all of whom try to report on the feasibility of their particular initiative. (I do recognize that the authors accurately and appropriately reflect on this – albeit in just one sentence – on pg 19, line 32/33). But it seems that if this paper is going to provide something meaningful to the reader, it should be trying to unpack these results to understand what aspects of the clinic environment, or consumer demand Integra failed to address and conversely - why

	in the clinics where integrated service delivery persisted, it did so. Minor Revisions • Pg 6, line 14: suggest changing 'hindering' to 'hinders' • Pg 6, line 21: suggest inserting 'services' or 'care and treatment' after 'HIV' • Pg 7, line 42: is 'regions' the technical term? Are they districts? Provinces?

REVIEWER	Mary Lou Lindegren Vanderbilt University USA
REVIEW RETURNED	09-Oct-2013

GENERAL COMMENTS	To the editors: This paper needs revisions. There is no way to fix the systematically different groups of facilities in the intervention and comparison clinics or the timing of the assessments being inconsistent over the course of the study. The results demonstrate heterogeneity across clinics, both intervention and comparison. Without contextual data it is hard to understand what this means in terms of capacity to integrate. There were no references to validated and reliable methodology provided for this client flow approach. However, if all the limitations are acknowledged, comments addressed, and the conclusions qualified this can contribute to the literature concerning feasibility of integration of services and challenges to data collection to measure integration. Recommendation: Revise and reconsider This paper assesses the integration of HIV and MCH services in 8 public sector facilities in Swaziland through a client flow assessment, where an intervention to strengthen integration of HIV into postnatal care (the Integra Initiative) was underway at 4 of the 8 sites. Clinics were purposefully selected for the intervention sites. Client flow assessments conducted over 5 days annually in 2009, 2010, and again in 2012 to determine the number of HIV/STI services that were received with MCH services in the same visit and in what combination and to determine whether receipt of integrated services differs between facilities which did and did not receive the Integra intervention. Client flow assessments were a novel approach to measuring the degree of integration. This study addresses an important question and is embedded into a broader unique and innovative study to look at integration of services (Integra Initative). However, due to systematic differences in clinics between intervention and comparison and the change in timing of yearly client flow assessment in 2010, they could not provide any definitive conclusion regarding trends in the implementation of integrated services across clinics nor compare intervention with comparison clinics. They were able to demonstrate the feasibility of integrated services, and the heterogeneity of implementation and sustainability. Abstract Unclear what the intervention was exactly "activities and resources to strengthen integrations o HIV services into post-natal care"—can you be any more specific Can you provide data to support the results, such as the proportion of different types of visits where services were integrated (ANC, child health, PMC, cervical screening);

	Methods Can you provide any more information comparing characteristics of intervention and comparison clinics. (e.g., level of care-primary, secondary, tertiary); From the data in Table 1 they appear quite different in terms of setting (comparison clinics mostly rural and intervention clinics mostly urban), annual client load (lower volume--6959-28,202 in comparison clinics compared to 9974-65,794 in intervention clinics. Were the comparison clinics matched by size to the intervention clinics? From the text the only criteria is distance from the intervention sites and no current provision of integrated services. I am concerned you have comparison clinics that are very different than intervention clinics. Also, the baseline service provision of both any MCH and any HIV/STI services looks quite different across comparison and intervention clinics (supplementary table 1). At baseline provision of HIV services ranges from 9.7% to 43.6% in intervention clinics to 13.6% to 50.7% in comparison clinics. Client flow assessments were a novel approach to measuring the receipt of integrated servics, however, no references to validated methodology for this approach were provided. Can you provide references to how client flow reliability and validity has been evaluated. The baseline data were collected after the intervention had begun to be implemented, can you comment on the implications of this regarding looking at changes over time and comparisons to non intervention clinics. Were there any power calculations conducted to assess what sample size of visits was needed in each clinic to detect a difference over time or compared to intervention clinics? In some clinics the sample size was quite variable over the years. In one facility sample size in one year was very small compared to other years (facility H sampled 169 visits in 2009, 35 visits in 2010 and 287 visits in 2012). One of the clinics where the proportion of integrated visits decreased in 2010 had drastic drop in number of visits ascertained as well (Facility B N=855 in 2009, N=263 in 2010, and N=408 in 2012) Can you comment on how this variability could have affected the results? It is hard to interpret trends in the percentage when the denominator is changing so much from year to year in the same clinic. Results In Table 2, at baseline the proportion of visits where integrative services were received was higher 25.4% (range 10.9%49.3%) in comparison clinics than intervention clinics (16.1%, range 9.2%-32.7%) The number of visits ascertained and the proportion with integrated services was extremely variable across clinics and within some clinics (Facility G, Facility F), making one wonder about the methodology and the comparability of one year to the next in ascertaining visits and receipt of services. It would be helpful to look at the number (%) of any HIV services received over time compared to the number (%) of integrated HIV services over time and in each clinic to determine if there vertical programs that had separate interventions that impacted receipt of integrated services. Figure 1 is a nice illustration, but there is no scale provided to ascertain what the size of the bubble corresponds to, this should be a standard range. Also, it appears that receipt of integrated SRH-

	HIV services was much greater than MCH-HIV services, can you comment on these data. Discussion A large limitation of study is that the comparison sites could not be matched with similar size intervention sites, which authors acknowledge in the limitations section, which resulted in intervention facilities that are systematically different (largely urban) than comparison facilities (mostly rural). Additionally, client flow assessments were not conducted during the same week in 2010 as in 2009, in 2010 most assessments were delayed until the week before Christmas resulting in smaller number of clients and likely different range of services provided. The methods, timing of the assessments, and deviations from the protocol should be noted in the methods of the paper as well. The authors note that five days may not be long enough, yet this assessment was resource intensive. We would benefit from references to standard approaches to client flow assessment to enhance reliability and validity of data.

VERSION 1 – AUTHOR RESPONSE

Reviewer Name SM Topp
Institution and Country Centre for Infectious Disease Research in Zambia
Please state any competing interests or state 'None declared': None delcared

The authors have clearly put considerable effort into this manuscript and I applaud their hard work. However, I think the conclusions drawn from the data are frequently overstated, and the manuscript presented requires considerable reworking.

Compulsory Revisions
Introduction:
• Last sentence of para 2, pg 6: the authors refer to 'process evaluations' – but only reference one paper. Is there more than one?
[Note this is a reference to a literature review, in which the findings of numerous evaluations are synthesised.]
• Para 3, pg 6: the authors present findings from a systematic review which concludes that integration of services is 'feasible'. What this does not account for is that the studies reviewed would have shown integration was feasible under certain circumstances. Such findings are never an absolute. When using such studies as background, it is important to qualify the use of such evidence with a demonstration of your understanding that programmatic 'success' is always context specific.
[We have qualified this statement by adding 'under certain circumstances'.]
• Para 3, pg 6: the second half of this para present the evidence with a little more nuance and qualification; remember, integration can yield benefit, but it is not an automatic outcome of any integration intervention. Previous experience implementing integrated programs elsewhere is not a guarantee. Integration requires a series of complex actions to take place and if any one goes wrong, then it may not have the intended outcome.
• General intro: the authors don't define what integration means in the context of your program. Integration can mean different things to different people; for the sake of clarity and precision the authors need to outline what this term means in your setting.
[We have now added "The specific models of integration, including the use of post-natal care as an

entry point for HIV/STI services in Swaziland, and their hypothesised benefits for clients and health care efficiency are detailed in the Integra study protocol", with a reference to the published protocol. Also, Outcome 1, defined in the Methods, represents the definition of integration used in this analysis.]

Methods:
• Pg 7, line 52: how did the authors determine that the control facilities had no integrated HIV-PNC services? This should be clarified in relation to the above suggestion to define what integration means in this setting.
[We have now added, "as determined by discussions with the Ministry of Health and site visits by Population Council."]

• Pg 8, line 5: Integra project was designed to strengthen and maintain the provision of integrated HIV and PNC services – but the authors don't specify what the anticipated benefits of integrated (versus non-integrated) service delivery would be. This is important to understanding the study's own objectives.
[We have now added to the Introduction, "The specific models of integration, including the use of post-natal care as an entry point for HIV/STI services in Swaziland, and their hypothesised benefits for clients and health care efficiency are detailed in the Integra study protocol", with a reference to the published protocol.]
• Pg 10, The very different nature of the control and intervention clinics calls into the question the validity of this comparison. Since the main measure was a facility-specific process (i.e. integrated services) I don't really understand the concern about contamination?
[We have preserved the a priori design to compare intervention and comparison facilities, but addressed the limitations of this comparison in the Discussion and presented within-facility change over time.]
[While the main measure in this analysis is facility-specific, individual level measures are being assessed in analyses of other data collected, from the same facilities, as part of the Integra Initiative (e.g., a client cohort study, research with providers, etc). Nevertheless, we have removed reference to the risk of contamination (in the Discussion/Limitations) given it is less relevant to this analysis and may create confusion.]
o Moreover – it would appear that there were other forms of 'integration' interventions even in the control facilities
[Given there is no true 'control' in a real-world setting, we have used the term 'comparison' rather than 'control' throughout.]

o The authors need to more clearly outline the contextual / service differences in these facilities – e.g. what service units did each one have; how many professional staff; how many lay staff; what was the catchment population and catchment area; how well supplied were they…all these things would be enormously important to understanding whether and how well services become integrated.
[We have added a supplementary table with additional characteristics about each facility, from Integra's health facility assessments (new Supplementary Table 1), and refer to the considerable differences in facility size, client volumes and staffing levels, in the Discussion/Limitations.]

• pg 15, lines 33/36: use of the term 'disappeared' is disingenuous. It would seem that healthcare providers simply ceased to deliver integrated services. The outstanding question would appear to be, why? (possible reasons: they felt it was outside their line of duty; they felt it was too much work; they experienced resistance from clients; they didn't have appropriate supplies; they forgot / they were called away or trained to do it differently /new staff transferred in and handed received the orientation).
• Simliar questions apply to the findings under sub-heading receipt of HIV counselling.
[These sentences have been edited to replace the word 'disappeared'.]

[The objectives of this analysis are to determine whether clients received integrate services over time, and in what combinations of services. We aim to describe the levels and patterns of use in more detail than possible from routine health information systems or previous studies. We agree it is also important to understand why levels and patterns of integration change, and are exploring this in depth with complementary data sources, but feel such a causal analysis is beyond the scope and length of this paper's objectives.]

Discussion:
• The finding that there was no evidence of increased integration in the intervention clinics appears valid – but I find little additional value to the comparison with control clinics
[As we mention above, we have preserved the a priori design to compare intervention and comparison facilities, and felt it was important to include these results for transparency. We also address the limitations of this comparison in the Discussion and present within-facility change over time given the heterogeneity.]

• The absence of any qualitative (or anecdotal programmatic) data with which to meaningfully interpret the quantitative findings is a significant limitation. This means that the authors rely on conjecture to understand the relative lack of integration of service delivery and lack of meaningful patterns in integration across program vs. control clinics.
[We have not sought to explain causes of the findings, as understanding the reasons for differences between facilities and changes over time is a substantial, additional analysis (understanding dose, fidelity and quality of the intervention is underway with additional data sources; we now refer to this follow-on work in the Discussion). We maintain that understanding the levels and detailed patterns of integration in public sector facilities is a valuable contribution to the literature and starting point for in-depth causal analyses.]

• Pg 18, line 5/6: the authors suggest that 'HIV counselling…has an important role to play in scaling up integration'— in doing so they seem to suggest that integration is an inherent good. This is manifestly not the case. Integration is a means to an end – whether that be more efficient or less stigmatizing health care. Much more care needs to be taken by the authors in considering the implications of their results. I would like to understand if they have prima facae evidence of integration providing programmatic or clinical benefit in the Swazi context – if such evidence does exist, then this must be much more clearly presented, either as part of a literature review or in the results.
[We have removed the word 'important' from this sentence. Our intention is to show that, where integration improved (for better or for worse), HIV counselling played a role in this increase.]

• Pg 18, lines 37-48: again the authors seem to suggest that the effectiveness of the Integra initiative is a given – and that it is the clinics or context in which they are working that needs to change to demonstrate this effectiveness. The results would suggest that the Integra initiative failed to meaningfully address a range of programmatic, operational and broader contextual factors relating to the provision of integrated MCH services. THIS is a useful and meaningful (even if less palatable) finding in the context of many, similar programs, all of whom try to report on the feasibility of their particular initiative. (I do recognize that the authors accurately and appropriately reflect on this – albeit in just one sentence – on pg 19, line 32/33). But it seems that if this paper is going to provide something meaningful to the reader, it should be trying to unpack these results to understand what aspects of the clinic environment, or consumer demand Integra failed to address and conversely - why in the clinics where integrated service delivery persisted, it did so.
[We have added the word 'potential' before 'effectiveness of the Integra Initiative'. As mentioned above, we are conducting more detailed case studies to understand the absence of an intervention effect (which we acknowledge directly in the Results and Discussion), and other interesting questions this analysis has raised.]

Minor Revisions
• Pg 6, line 14: suggest changing ‘hindering’ to ‘hinders’
[This has been changed.]
• Pg 6, line 21: suggest inserting ‘services’ or ‘care and treatment’ after ‘HIV’
[This has been changed.]
• Pg 7, line 42: is ‘regions’ the technical term? Are they districts? Provinces?
[Yes, they are called regions.]

Reviewer Name Mary Lou Lindegren
Institution and Country Vanderbilt University
USA
Please state any competing interests or state ‘None declared’: None declared

This paper needs revisions. There is no way to fix the systematically different groups of facilities in the intervention and comparison clinics or the timing of the assessments being inconsistent over the course of the study. The results demonstrate heterogeneity across clinics, both intervention and comparison. Without contextual data it is hard to understand what this means in terms of capacity to integrate. There were no references to validated and reliable methodology provided for this client flow approach. However, if all the limitations are acknowledged, comments addressed, and the conclusions qualified this can contribute to the literature concerning feasibility of integration of services and challenges to data collection to measure integration.

This paper assesses the integration of HIV and MCH services in 8 public sector facilities in Swaziland through a client flow assessment, where an intervention to strengthen integration of HIV into postnatal care (the Integra Initiative) was underway at 4 of the 8 sites. Clinics were purposefully selected for the intervention sites. Client flow assessments conducted over 5 days annually in 2009, 2010, and again in 2012 to determine the number of HIV/STI services that were received with MCH services in the same visit and in what combination and to determine whether receipt of integrated services differs between facilities which did and did not receive the Integra intervention. Client flow assessments were a novel approach to measuring the degree of integration. This study addresses an important question and is embedded into a broader unique and innovative study to look at integration of services (Integra Initative). However, due to systematic differences in clinics between intervention and comparison and the change in timing of yearly client flow assessment in 2010, they could not provide any definitive conclusion regarding trends in the implementation of integrated services across clinics nor compare intervention with comparison clinics. They were able to demonstrate the feasibility of integrated services, and the heterogeneity of implementation and sustainability.
[We appreciate these comments, and the reviewer’s consideration of the study’s strengths and limitations.]

Abstract
Unclear what the intervention was exactly “activities and resources to strengthen integrations o HIV services into post-natal care”—can you be any more specific
Can you provide data to support the results, such as the proportion of different types of visits where services were integrated (ANC, child health, PMC, cervical screening);
[We will defer to the Editor whether these details can be added to the Abstract, given the Abstract currently has the maximum word limit of 300.]

Methods
Can you provide any more information comparing characteristics of intervention and comparison clinics. (e.g., level of care-primary, secondary, tertiary); From the data in Table 1 they appear quite different in terms of setting (comparison clinics mostly rural and intervention clinics mostly urban),

annual client load (lower volume--6959-28,202 in comparison clinics compared to 9974-65,794 in intervention clinics. Were the comparison clinics matched by size to the intervention clinics?
[We have added a supplementary table with additional characteristics about each facility, from Integra's health facility assessments (new Supplementary Table 1), and refer to the considerable differences in facility size, client volumes and staffing levels, in the Discussion/Limitations.]
[It was not possible to match facilities by size, given Swaziland is a small country with few overall facilities. (This is described in the Methods and also included as a limitation in the Discussion.)]

From the text the only criteria is distance from the intervention sites and no current provision of integrated services. I am concerned you have comparison clinics that are very different than intervention clinics. Also, the baseline service provision of both any MCH and any HIV/STI services looks quite different across comparison and intervention clinics (supplementary table 1). At baseline provision of HIV services ranges from 9.7% to 43.6% in intervention clinics to 13.6% to 50.7% in comparison clinics.
[We have preserved the a priori design to compare intervention and comparison facilities, and felt it was important to include these results for transparency. However, we address the limitations of this comparison in the Discussion and present within-facility change over time – an approach we felt was more informative given the heterogeneity.]

Client flow assessments were a novel approach to measuring the receipt of integrated servics, however, no references to validated methodology for this approach were provided. Can you provide references to how client flow reliability and validity has been evaluated.
[We have now included new references in the Introduction, related to the development of the 'patient flow analysis' by the CDC (upon which the client flow methodology was based) and evaluation of the method in family planning clinics in Kenya.]
[We also cite in the Discussion lessons from past evaluations of this methodology.]
The baseline data were collected after the intervention had begun to be implemented, can you comment on the implications of this regarding looking at changes over time and comparisons to non intervention clinics.
[The baseline data were collected before the intervention was fully implemented in any facility, and levels of integration were lower in intervention than comparison sites at baseline.]
Were there any power calculations conducted to assess what sample size of visits was needed in each clinic to detect a difference over time or compared to intervention clinics?
[There was no sampling strategy as this element of the Integra Initiative was designed as a census of all clients seen over a fixed period (as per the CDC 'patient flow analysis' method; Lynam 1994). Our focus, therefore, was to construct confidence intervals around the levels of integration observed, where the width of the confidence intervals reflect the precision of our estimates given the data available. Whereas the CDC method recommends tracking clients over one day, we extended tracking to five days to maximise learning.]
In some clinics the sample size was quite variable over the years. In one facility sample size in one year was very small compared to other years (facility H sampled 169 visits in 2009, 35 visits in 2010 and 287 visits in 2012). One of the clinics where the proportion of integrated visits decreased in 2010 had drastic drop in number of visits ascertained as well (Facility B N=855 in 2009, N=263 in 2010, and N=408 in 2012) Can you comment on how this variability could have affected the results? It is hard to interpret trends in the percentage when the denominator is changing so much from year to year in the same clinic.
[In the Discussion, we acknowledge the smaller number of clients in most facilities in 2010, which we attribute to the Christmas season. We also acknowledge that the timing may have affected the range of services provided and different patterns of integration. Given these fluctuations (and the reality that client volume will rarely be constant, now explicitly addressed in the Discussion with reference to past analyses of patient flow), we suggest that it may be more informative to monitor over a longer period than five days, or ideally through integrated routine health information systems.]

Results
In Table 2, at baseline the proportion of visits where integrative services were received was higher 25.4% (range 10.9%49.3%) in comparison clinics than intervention clinics (16.1%, range 9.2%-32.7%)

The number of visits ascertained and the proportion with integrated services was extremely variable across clinics and within some clinics (Facility G, Facility F), making one wonder about the methodology and the comparability of one year to the next in ascertaining visits and receipt of services. It would be helpful to look at the number (%) of any HIV services received over time compared to the number (%) of integrated HIV services over time and in each clinic to determine if there vertical programs that had separate interventions that impacted receipt of integrated services.
[We did not assess this formally, as it was not a key objective, however, Supplementary Table 2 (previously 1) shows the proportion of visits in which any HIV/STI services were received and specifically HIV treatment. In the facilities in which HIV treatment rose substantially (C, G, H), integrated services declined more than in other facilities, suggesting that integration may be impacted by vertical programmes (a point made in the Discussion). Also, unlike HIV counselling, HIV treatment was more often provided alone than in combination with an MCH service.]

Figure 1 is a nice illustration, but there is no scale provided to ascertain what the size of the bubble corresponds to, this should be a standard range. Also, it appears that receipt of integrated SRH-HIV services was much greater than MCH-HIV services, can you comment on these data.
[The footnote is included to provide a measure of scale, noting that the largest bubble represents 49% of all visits.]
[Thank you for identifying an error in our labelling. The label for the final column has been corrected to 'MCH' (previously 'SRH' creating confusion), and an updated Figure 1 has been uploaded.]

Discussion

A large limitation of study is that the comparison sites could not be matched with similar size intervention sites, which authors acknowledge in the limitations section, which resulted in intervention facilities that are systematically different (largely urban) than comparison facilities (mostly rural). Additionally, client flow assessments were not conducted during the same week in 2010 as in 2009, in 2010 most assessments were delayed until the week before Christmas resulting in smaller number of clients and likely different range of services provided.

The methods, timing of the assessments, and deviations from the protocol should be noted in the methods of the paper as well.
[We have included text about logistical constraints on intended timing, within the Methods (end of Data Collection section).]

The authors note that five days may not be long enough, yet this assessment was resource intensive. We would benefit from references to standard approaches to client flow assessment to enhance reliability and validity of data.
[We agree and have added references to new citations 13, 14, 15 in the Introduction and Discussion.]

VERSION 2 – REVIEW

REVIEWER	Mary Lou Lindegren Vanderbilt University School of Medicine
REVIEW RETURNED	12-Dec-2013

GENERAL COMMENTS	Thank you for the excellent revisions. I think this paper highlights important challenges in integration and measurement of integration. I have several concerns that I do not feel were adequately addressed in the authors comments. See below Introduction and methods The authors need to define in the paper what integration means in this setting, referring to protocol as noted below is not specific enough. Likewise, clarification of what it means to have no integrated services for the control facilities. There are definitions in the literature that could be cited, or outlining exactly how it was defined in the protocol is important. Results and Discussion There is no way to fix the systematically different groups of facilities in the intervention and comparison clinics. The results demonstrate heterogeneity across clinics, both intervention and comparison. Without contextual data it is hard to understand what this means in terms of capacity to integrate. I remain concerned you have comparison clinics that are very different than intervention clinics. Also, the baseline service provision of both any MCH and any HIV/STI services looks quite different across comparison and intervention clinics (supplementary table 2). At baseline provision of HIV services ranges from 9.7% to 43.6% in intervention clinics to 13.6% to 50.7% in comparison clinics. I think as a result the approach to present within-facility change over time is more informative and should be the focus rather than the comparison of intervention and control facilities. Thank you for including supplementary tables outlining the characteristics of clinics in terms of staffing, etc. Thank you for including references on client flow assessments. The deviations from the protocol should be noted in the methods of the paper.

VERSION 2 – AUTHOR RESPONSE

Introduction and methods
The authors need to define in the paper what integration means in this setting, referring to protocol as noted below is not specific enough. Likewise, clarification of what it means to have no integrated services for the control facilities. There are definitions in the literature that could be cited, or outlining exactly how it was defined in the protocol is important.

[We have now added more detail about the definition of integration used across the Integra study and specifically in the Swaziland context (in the Introduction , p7). With regard to the control facilities, we have now added the year in which determinations were made together with the Ministry of Health and site visits that there was no ‘current’ provision of HIV-PNC integration in 2008 (p8). This was one year

before the client flow assessments started, and based on discussions rather than a precise tool like the client flow assessment. That integration was actually identified within the comparison facilities through the client flow assessments in 2009 is another reason we feel a focus on facility-specific rather than by-group analyses is appropriate. We have added this point to the Discussion on page 20.]

Results and Discussion
There is no way to fix the systematically different groups of facilities in the intervention and comparison clinics. The results demonstrate heterogeneity across clinics, both intervention and comparison. Without contextual data it is hard to understand what this means in terms of capacity to integrate. I remain concerned you have comparison clinics that are very different than intervention clinics. Also, the baseline service provision of both any MCH and any HIV/STI services looks quite different across comparison and intervention clinics (supplementary table 2). At baseline provision of HIV services ranges from 9.7% to 43.6% in intervention clinics to 13.6% to 50.7% in comparison clinics. I think as a result the approach to present within-facility change over time is more informative and should be the focus rather than the comparison of intervention and control facilities.

[We feel that within-facility change is already the focus of the analysis and results and discussion. As explained in our previous response, we retained the comparison by design group for transparency (to be consistent with the protocol) but feel it receives relatively little attention in the text and tables (only one column in Table 2), compared to facility-specific results, and its limitations are detailed in the Discussion.]

Thank you for including supplementary tables outlining the characteristics of clinics in terms of staffing, etc.
Thank you for including references on client flow assessments.

The deviations from the protocol should be noted in the methods of the paper.

[Deviations to the protocol are summarised in the final paragraph of the Methods, p9. In addition, we have now added that in some facilities, client flow assessments were conducted for more than the five days intended. We have restricted this analysis to data collected on the first Monday through Friday, to preserve the original protocol design.]