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](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. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

<table>
<thead>
<tr>
<th>TITLE (PROVISIONAL)</th>
<th>Developing Objective Metrics for Unit Staffing (DOMUS) study</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTHORS</td>
<td>Siddiqui, Iram; Whittingham, Beverley; Meadowcroft, Karen; Richardson, Mark; Cooper, Jason; Belcher, John; Morris, Eddie; Ismail, Khaled</td>
</tr>
</tbody>
</table>

VERSION 1 - REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Jane Sandall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>King's College, London.</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>11-May-2014</td>
</tr>
</tbody>
</table>

GENERAL COMMENTS

The derivation of staffing tools and escalation policies to ensure safe staffing on labour wards is an important issue, particularly in relation to capturing real time fluctuations in demand.

Such tools need to be easy to use, valid, reliable and effective, in terms of whether tools do trigger escalation, and whether escalation policies work in practice.

The aim of this study was to assess the feasibility of using a scoring system to contemporaneously assess the required staffing level based on demand and use this to determine delivery suite escalation level, and utilise the information generated regarding clinical activity (Demand) and staffing levels (Capacity) to generate unit-specific calculation for the actual number of midwifery staff required.

The paper reports results on an audit/quality improvement project that was conducted in one maternity provider over a 12 month period in 2009-2010, thus the paper is drawing on data from five years ago. I think this is a limitation that should be acknowledged. It would also be helpful to have an update from the authors.

Based on a dependency score (Demand) and the number of midwifery staff available (Capacity), an escalation level was determined for each shift. The 80th percentile of the demand was used to determine optimal capacity, and justification has been provided for this decision.

However, details of the derivation and validation of the demand and capacity score needs more detail. The score system is based on Birthrate plus, a reference for which needs to be provided, and in addition, it is important that the reader is provided an overview of the current evidence base of Birthrate plus.

The scoring system with reference to Figure 1 needs to be explained to readers who are not familiar with Birthrate plus.
The authors report that a comparison of midwifery vs investigator coding of level of dependency was conducted. It would be important to include more details and results regarding inter-rater variability regarding the level of agreement and how the findings inform the instructions for future coding and escalation trigger levels in the paper.

Such a tool needs to be sensitive enough to detect variation, but not oversensitive. There is a substantial literature on floor and ceiling levels and how outside factors influence coding which would add to this paper and inform the author’s analysis.

It would be helpful to the reader for the authors to provide more information regarding what dependency scores mean in reality (ie an average of 7). The authors state that the 80th percentile of the demand variation was calculated to be 11. What does this mean in real life staffing terms?

A total of 1160 of the potential 1568 (74%) scoring sheets were completed. The authors hypothesise that non-completion would occur at the busiest times and thus missing opportunities for escalation. With a 26% non-completion rate, it would be helpful for the authors to provide more information regarding reasons for non-completion.

More detail needs to be provided regarding what actions were intended following escalation, and whether this actually happened. Figure 3 shows escalation during the study period, but it is unclear what the count on the y axis is of, when % is being used for the columns. Most importantly, what was the response to escalation and how often was it not possible to respond.

It is unclear whether this is research or a report of a quality improvement audit, but it is important to state whether ethical approval was in place.

In summary, this paper reports early piloting of the use of a tool to measure and respond to fluctuations in demand and capacity on labour wards. However, more information needs to be provided regarding validation (if conducted by the team) or more work needs to be done regarding feasibility, barrier to completion and ease of use, inter-rater reliability, trigger levels and response before such a tool could be used in other settings.

REVIEWER
Tim Hillard
Poole Hospital NHS Foundation Trust
Chairman RCOG/NHS England Patient Safety Expert Group

REVIEW RETURNED 15-May-2014

GENERAL COMMENTS
This paper explores the use of a new tool for realtime assessment of midwifery staffing levels on the delivery suite. Whilst there are several existing tools for this already in use this new version modifies the most widely used one, BR+, by increasing the scenarios included in the dependency assessment and increasing the time points at which the tool is applied from 2 - 4/24hrs. The paper is well argued and presented and has the potential to be
a useful clinical tool for many units. The limitations of the study are laid out and discussed appropriately. My main concern at a practical level is the setting of the optimum level at 80%. Whilst the rationale for this target level is well argued, it is a considerable jump from the 50% level currently employed by many trusts. The resource implications of this are considerable. It may be that by moving staff around the units to meet the demands the costs could be restricted but it would be helpful if the paper gave some indication as to how this could be achieved. It would also be useful to know if further evaluation of this tool in a larger, and perhaps more diverse, number of units is intended.

The staffing assessment is solely reliant on number of midwives. Has there been any attempt to widen the scope of this tool to reflect capacity based on midwifery skill mix or banding? For instance a shift with 5 experienced midwives may potentially have greater capacity than one with 7 newly qualified midwives. Other factors that may also affect capacity modelling include the level of junior doctor cover (experienced or new) and number and skill mix of support staff.

---

**REVIEWER**

SIMON CUNNINGHAM
LEIGHTON HOSPITAL
CREWE
CHESHIRE
UNITED KINGDOM

**REVIEW RETURNED** 16-May-2014

**GENERAL COMMENTS**

The 21st century obstetric population demonstrates increased levels of risk and complexity with hitherto unmatched levels of obesity, maternal age, chronic disease and congenital disease at a time when continuity of care is challenged by new working patterns. Traditional staffing models are being exposed and services are facing accusation of understaffing on a regular basis. The challenge has become to staff for activity rather than capacity. This model accepts the need for an increased number of staff but in response to an increased level of activity seen within a highly dynamic environment such as labour ward.

The methods within the paper have attempted to chart changes within the level of activity and staff and establish if the unit is functioning optimally rather than continually overextended and likely to face institutional exhaustion.

It takes established tools such as the birth rate plus to assess the level of activity and generate the staffing need for the delivery suite in a dynamic fashion. This is then compared against the actual staffing levels.

It clearly demonstrates the need to monitor levels of activity and staffing and to be able to react efficiently when taking escalation measures to ensure the unit remains safe. This has hitherto been carried out in a relatively subjective fashion. The value of the study is that it demonstrates the application and validation of a tool that provide an objective, transparent and clearly defined overview of the situation. As such it provides the basis for next actions to be carried out faster and appropriately. The main aim of the study was to assess the feasibility of using a modified BR+ scoring approach to
determine the level of DS escalation based on demand and staffing levels. It has demonstrated that it is feasible to create model that generates consistent information on which to base decisions.

The study was carried out at the same time that the birthrate acuity intrapartum software system was developed using similar models. To the best of my knowledge no publication has occurred using the birthrate acuity software. This study takes the principles used within the software design and applies them in a realtime prospective context.

There are several limitations

The frequency of monitoring
Failure to enter 26% of scoring sheets
The birthrate methodology was changed to include a class for induced labours
The lack of quality control on scoring sheets to define inter observer variation or bias in scoring the sheets.

Monitoring frequency
The authors acknowledge this as a limitation and demonstrate progress from an initial design that monitored twice per day. Indeed the birth rate intrapartum acuity software has different frequencies of monitoring available for units starting at every two hours. Whilst the labour ward environment is highly dynamic prospectively monitoring every 6 hours is more achievable in this context. The data suggests that tool was able to catch most of the variation at this frequency.

Missing data
The authors acknowledge this. It appears to occur most early morning. Data is limited in defining if this has biased the results or there was a specific cause. I would suggest that the main cause is that the study relies on the enthusiasm and cooperation of staff involved. Simply at four in the morning there is less peer pressure and supervision than during daylight hours. I do not think that it suggests a consistently high level of activity. Given the volume of data available the authors could have analysed that time period across the year and demonstrated if it was consistently busier.

Modification of the birthrate plus tool
Modifying the birth rate plus methodology
I would suggest that the authors understood a limitation in their existing methodology. They changed design to generate a more accurate dataset. As such the results are more likely to reflect the true nature of activity then if the study had continued with no modifications. With successive versions of the birth rate intrapartum acuity software have now also incorporated the impact of induced labours within the activity levels.

Intra and Interobserver variation
A 6 month pilot was carried out prior to the study and this provided an important opportunity for staff to acclimatise to the methodology. Given the study was carried out over a year there would have been an opportunity to demonstrate the level of intra and inter observer variation for staff recording levels of activity. The study is however a feasibility project looking to apply the tool to real life dynamics. With later work it would be extremely useful to establish the influence such variation has on scoring and activation of escalation measures.
I think that a breakdown of the reasons for unit closure and the influence staffing had on this would have been useful. It could have been placed alongside the scores obtained from the tool to demonstrate emerging pressures, measures taken and an opinion as to whether diversion could have been prevented. However this represents the next level of work.

In the post Francis NHS where we understand how institutional exhaustion saps empathy and compassion and where NICE are beginning to provide guidance on appropriate levels of staff the paper represents an important first step in defining safe staffing levels.

While the references are up to date, given this represents an extremely topical debate I would suggest adding some from the last two years.

**VERSION 1 – AUTHOR RESPONSE**

Reviewer 1

1A) The paper reports results on an audit/quality improvement project that was conducted in one maternity provider over a 12 month period in 2009-2010, thus the paper is drawing on data from five years ago. I think this is a limitation that should be acknowledged. It would also be helpful to have an update from the authors.

Response

We thank the reviewer for the valid comment about the time that elapsed between completion of the study and its submission for publication however this has happened because of unforeseen circumstances relating to job relocation. With regards to an “update”, this is not feasible because data was collected during the pilot phase of testing this tool prior to its full incorporation in routine practice in the unit. To our knowledge, once the new escalation policy was included as part of routine practice and the author (IS) responsible for data handling completed her out of programme placement detailed data storage was not required.

1B) Based on a dependency score (Demand) and the number of midwifery staff available (Capacity), an escalation level was determined for each shift. The 80th percentile of the demand was used to determine optimal capacity, and justification has been provided for this decision. However, details of the derivation and validation of the demand and capacity score needs more detail. The score system is based on Birthrate plus, a reference for which needs to be provided, and in addition, it is important that the reader is provided an overview of the current evidence base of Birthrate plus.

Response

Further to the reviewer’s comment we have add the following text to the introduction section introducing the concept behind the development of Birthrate plus scoring and how it relates to capacity and demand. We have also added the relevant references. “Intrapartum scoring based on mother and baby care needs was originally developed in 1986 when the score reflected the level of clinical complexity and hence demand. The methodology was applied in several hundred maternity services and extensive validation of the dependency scores was done. More recently, Birth Rate Plus (BR+) was developed and validated as a real-time objective measure of the number of midwives required to provide a high quality maternity service at any point in time on delivery suite (DS). The
principle is based on the knowledge that women with higher care needs require more than 1:1 midwife time to care for them whilst on labour ward[7]. Hence, the total BR+ score at anytime on labour ward determines the minimum number of midwives expected to be available to provide intrapartum care at that time point."

1C) The score system is based on Birthrate plus, a reference for which needs to be provided, and in addition, it is important that the reader is provided an overview of the current evidence base of Birthrate plus.

Response
The following references in support of the development and validation of the Birthrate Plus tool were added to the manuscript.


1D) The scoring system with reference to Figure 1 needs to be explained to readers who are not familiar with Birthrate plus.

Response
We thank the reviewer for highlighting this area which seems to be vague in our original manuscript. We believe that this should be clearer now following the amendment we made to the paragraph explaining the concept behind the development of BR+ (see 1D). We have also added the following text to the discussion section: “However, IOL is ......... in BR+ (an intrapartum scoring system it only accounts for women in labour), ...... new category was added...”

1E) The authors report that a comparison of midwifery vs investigator coding of level of dependency was conducted. It would be important to include more details and results regarding inter-rater variability regarding the level of agreement and how the findings inform the instructions for future coding and escalation trigger levels in the paper.

Response
Thank you for giving us the opportunity to clarify this important issue. The comparison between investigator and midwife in charge coding was related to the final documentation of level of dependency and not related to scoring of individual cases (which was not performed). Therefore, this variability would not have impacted on the capacity / demand analysis we conducted. We believe we have explained this and possible reasons for these differences in the discussion section, Paragraph 4.

1F) Such a tool needs to be sensitive enough to detect variation, but not oversensitive. There is a substantial literature on floor and ceiling levels and how outside factors influence coding which would add to this paper and inform the author’s analysis.

Response
We totally agree with the reviewer, however, this aspect is related to validation of BR+ as stated at the end of the introduction section "The main aims of the DOMUS study were to assess the feasibility of
using a modified BR+ scoring approach to determine the level of DS escalation and utilise the information generated regarding clinical activity (Demand) and staffing levels (Capacity) to generate unit-specific calculation for the actual number of midwifery staff required to deliver a high quality service using a validated capacity and demand model."

1G) It would be helpful to the reader for the authors to provide more information regarding what dependency scores mean in reality (i.e., an average of 7). The authors state that the 80th percentile of the demand variation was calculated to be 11. What does this mean in real life staffing terms?

Dependency scores calculated based on the demand (number of staff required to deliver optimal care for the number and case mix of women on labour ward at a point in time. The average value of this in our study was 7. Further to the reviewer’s comment we have clarified this in paragraph 3 in the introduction “In an attempt to reduce the subjectivity and determine the level of escalation of DS in a busy maternity unit, we conceived the idea of using a modified BR+ scoring system to contemporaneously assess the required staffing level (dependency score) based on demand and use this score to determine DS escalation level.” and the second paragraph in the methods section “Based on the dependency score determined by the number and clinical needs of women on DS (Demand) in relation to the number of midwifery staff available (Capacity), an escalation level was determined for each shift. A traffic light system was used to display the escalation”. Dependency scores are calculated real time. In contrast, the use of the 80th percentile of demand variation to determine levels of staff deployment are calculated based on data collected over a period of time. In this study it was 11 which means that 11 midwives would be required per shift to ensure that appropriate staffing levels are available to meet the demand in 80% of the times.

1H) A total of 1160 of the potential 1568 (74%) scoring sheets were completed. The authors hypothesise that non-completion would occur at the busiest times and thus missing opportunities for escalation. With a 26% non-completion rate, it would be helpful for the authors to provide more information regarding reasons for non-completion.

We agree with the reviewer that this information could have provided better insight into reasons for non-completion, however, we did not collect this information as part of this study. We have added a sentence to paragraph 5 of the discussion to highlight this limitation and its potential implications. “Unfortunately, we did not collect information regarding reasons for not completing a score sheet hence our proposed reasons for not completion are only speculative……”

1I) More detail needs to be provided regarding what actions were intended following escalation, and whether this actually happened. Figure 3 shows escalation during the study period, but it is unclear what the count on the y axis is of, when % is being used for the columns. Most importantly, what was the response to escalation and how often was it not possible to respond?

Response We have provided this information in paragraph 3 of the results sections. “The unit escalation level according to the scored sheets……”. Once an escalation level was decided based on the score sheet it was followed including escalation level 4. We have also included the number of occasions the unit was closed during the study period. However, it is imperative to read this in the context of the issue highlighted in Paragraph 4 of the discussion where a discrepancy was noted between what labour ward coordinators have determined as an escalation level to an independent assessor who retrospectively assessed escalation particularly in the extremes of the escalation grades (i.e., Green
and red). Figure 3 shows the percentage as well as the actual number of events.

1J) It is unclear whether this is research or a report of a quality improvement audit, but it is important to state whether ethical approval was in place.

Response
This study was undertaken as a quality improvement development for patient safety and staff development. Therefore, the score sheet and escalation policy were agreed by the Trust Board Clinical Governance Board prior to piloting. We added a sentence to the second paragraph of the methods section to reflect this: “Following agreement from the Trust clinical governance board, piloting the escalation policy was…….”

1K) In summary, this paper reports early piloting of the use of a tool to measure and respond to fluctuations in demand and capacity on labour wards. However, more information needs to be provided regarding validation (if conducted by the team) or more work needs to be done regarding feasibility, barrier to completion and ease of use, inter-rater reliability, trigger levels and response before such a tool could be used in other settings.

Response
We thank and agree with the reviewer. It is important to emphasize that we never suggested that this tool should be rolled out based on this study. To stress this point, we have amended the conclusion to reflect this: “This study highlights the feasibility of using a simple tool to determine the level of escalation on delivery suite based on an objective scoring system that takes into account capacity and demand. Moreover, this tool can be used to determine the appropriate staffing levels required for the safe and efficient service delivery on a labour ward. However, further work is required to assess stakeholders acceptability, inter-rater reliability and escalation trigger levels in different types of maternity units prior to wider use.”

Reviewer 2
2A) My main concern at a practical level is the setting of the optimum level at 80%. Whilst the rationale for this target level is well argued, it is a considerable jump from the 50% level currently employed by many trusts. The resource implications of this are considerable. It may be that by moving staff around the units to meet the demands the costs could be restricted but it would be helpful if the paper gave some indication as to how this could be achieved. It would also be useful to know if further evaluation of this tool in a larger, and perhaps more diverse, number of units is intended.

Response
We agree with the reviewer that staff deployment is the biggest challenge and the highest cost for any service delivery. However, the use of the 80th centile, as highlighted in our manuscript, seems the most efficient way of running a service on the long run. We recognise that the way different organisations achieve this will vary depending on the unit size. However, the real value of our tool is that it takes into account the case mix rather than just the number of cases. We believe that it is not within the remit of this study to suggest organisational measures of how to achieve optimal staffing levels.

2B) Other factors that may also affect capacity modelling include the level of junior doctor cover
Response
We agree with the reviewer and this will vary from one unit to another. However, Birthrate Plus only takes into account midwifery staffing using 1:1 midwifery care as the standard. Although the roles of midwives and obstetricians are complementary and integrated, the main care provided to women in labour is through midwifery staff.

Reviewer 3
3A) Monitoring frequency: The authors acknowledge this as a limitation and demonstrate progress from an initial design that monitored twice per day. Indeed the birth rate intrapartum acuity software has different frequencies of monitoring available for units starting at every two hours. Whilst the labour ward environment is highly dynamic prospectively monitoring every 6 hours is more achievable in this context. The data suggests that tool was able to catch most of the variation at this frequency.

Response
We agree with and thank the reviewer for this comment.

3B) Missing data: The authors acknowledge this. It appears to occur most early morning. Data is limited in defining if this has biased the results or there was a specific cause. I would suggest that the main cause is that the study relies on the enthusiasm and cooperation of staff involved. Simply at four in the morning there is less peer pressure and supervision than during daylight hours. I do not think that it suggests a consistently high level of activity.

Response
We agree that this is a possibility. However, we do not have evidence in our study to support this.

3C) Modification of the birthrate plus tool. Modifying the birth rate plus methodology. I would suggest that the authors understood a limitation in their existing methodology. They changed design to generate a more accurate dataset. As such the results are more likely to reflect the true nature of activity then if the study had continued with no modifications. With successive versions of the birth rate intrapartum acuity software have now also incorporated the impact of induced labours within the activity levels.

Response
We recognize the important point highlighted by the reviewer. As we explained in the introduction, the tool used in this study was developed before the NPSA intrapartum scorecard.

3D) I think that a breakdown of the reasons for unit closure and the influence staffing had on this would have been useful. It could have been place alongside the scores obtained from the tool to demonstrate emerging pressures, measures taken and an opinion as to whether diversion could have been prevented. However this represents the next level of work.

Response
We thank the reviewer for this valuable comment and will undoubtedly consider it when we further test this tool.

3E) While the references are up to date, given this represents an extremely topical debate I would suggest adding some from the last two years. The National Quality Board has published guidance on
nursing, midwifery and care staffing. It expects routine monitoring of shift to shift staffing levels and have systems such as e-rostering and escalation policies to be in place in order to quickly respond to changing demands. Ref: How to ensure the right people, with the right skills, are in the right place at the right time. A guide to nursing, midwifery and care staffing capacity and capability. NHS England Nov 2013.

Response
This was added to our reference list and included in the text.

VERSION 2 – REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Timothy Hillard</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poole Hospital NHS Trust</td>
</tr>
<tr>
<td></td>
<td>UK</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>01-Aug-2014</td>
</tr>
</tbody>
</table>

| GENERAL COMMENTS | The authors have satisfactorily addressed the initial reviewers comments and amended the paper accordingly. |

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Simon Cunningham</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Leighton Hospital</td>
</tr>
<tr>
<td></td>
<td>I am due to commence work at University Hospital Norths Staffs in October 2014</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>18-Jul-2014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GENERAL COMMENTS</th>
<th>Please can you incorporate comments from my first review.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>This paper clearly identifies the need to staff units objectively for activity rather than capacity and presents a possible method of doing this.</td>
</tr>
<tr>
<td></td>
<td>It demonstrates development of method, organisation learning, and quality control. There are no specific measurement relating to inter observer variation and intraobserver variation however these are alluded to within the text. The key question as to whether the DOMUS system is able to consistently and accurately record activity contemporaneous is answered.</td>
</tr>
<tr>
<td></td>
<td>I would suggest referencing recent publications from NICE re nursing staffing on hospital wards. This parallel details the need to staff for safety. The authors could demonstrate respective differences and contrast each system. Underlining both is a need to change organisation strategy to adapt to individual patient needs and avoid institutional exhaustion.</td>
</tr>
</tbody>
</table>
VERSION 2 – AUTHOR RESPONSE

We have added the suggested text and reference to the end of the discussion section as follows.

"Moreover, this study supports the view echoed in the recent NICE guideline relating to safe staffing which stresses that there is no single optimal nurse-to-patient ratio that can be applied across the board, however, each ward has to determine safe levels of staffing depending on their individual daily circumstances to ensure safe patient care. (31)"