National Audit Office

Report
by the National Audit Office

Department of Health

Modelling of maternity services in England

MARCH 2014

Our vision is to help the nation spend wisely.

Our public audit perspective helps Parliament hold government to account and improve public services.

The National Audit Office scrutinises public spending for Parliament and is independent of government. The Comptroller and Auditor General (C&AG), Amyas Morse, is an Officer of the House of Commons and leads the NAO, which employs some 860 staff. The C&AG certifies the accounts of all government departments and many other public sector bodies. He has statutory authority to examine and report to Parliament on whether departments and the bodies they fund have used their resources efficiently, effectively, and with economy. Our studies evaluate the value for money of public spending, nationally and locally. Our recommendations and reports on good practice help government improve public services, and our work led to audited savings of almost £1.2 billion in 2012.

Contents

Links to external websites were valid at the time of publication of this report. The National Audit Office is not responsible for the future validity of the links.

The National Audit Office team consisted of:
Michail Allios, Elena Cozzi,
Tom McBride and William Palmer.

This report can be found on the National Audit Office website at www.nao.org.uk

For further information about the National Audit Office please contact:

National Audit Office
Press Office
157–197 Buckingham Palace Road
Victoria
London
SW1W 9SP

Tel: 020 7798 7400

Enquiries: www.nao.org.uk/contact-us

Website: www.nao.org.uk

Twitter: @NAOorguk

Summary

Objectives

1 Having a baby is the most common reason for admission to hospital in England. There is limited published research evaluating whether maternity services can adapt to changes in demand and whether there is sufficient capacity to meet the Department of Health's objectives. We set out to examine these issues by modelling the maternity services at one NHS trust.

Methods

2 We used a 'discrete event simulation' to recreate (model) maternity services at a trust in England. The selected trust has an obstetric unit, and both an alongside midwifery unit (situated on the same site as the obstetric unit) and a freestanding midwifery unit. The regional NHS commissioning support service had forecast how the demand for maternity services could change in the next five years. However, it had not calculated the likely effect on resource requirements. To understand what pressures the services are likely to face, we used these forecasts to formulate scenarios for changes in demand and resource use. These included: increase in number and complexity of births and reduction in average length of stay. We also tested the resource implications of fully implementing the ambition for one-to-one care during established labour.

Results

3 The model suggests that bed numbers are sufficient and could probably absorb a relatively substantial increase in demand. However, the postnatal ward is shown to be the first ward to become full. The modelling results also provide evidence that providing one-to-one care during labour at all times is very challenging. An increased use of the midwifery-led unit, which is relatively underused, would ease resource pressures elsewhere in the trust. The modelling also confirms that reportedly achievable reductions in lengths of stay may provide a way of releasing substantial capacity.

Conclusion

4 There are issues with the data's quality, which mean that the model's outputs should be interpreted cautiously. This includes inconsistencies in the times recorded for mothers progressing to the different stages of labour. We did not intend to use the model to suggest a resource level for hospitals across England. Instead, we focused on a single trust to better understand the risks and challenges faced by providers. However, the work suggests that applying such methods more widely could help the Department of Health and the NHS to better understand how maternity services use resources.

Part One

Introduction

Overview

1.1 In November 2013, our report *Maternity services in England* examined the performance and management of maternity services.[1] We conducted fieldwork between January and September 2013, which included analysis of existing data and a survey of all NHS maternity services. One part of our methodology was a modelling exercise to provide an insight into local management issues and the resources required to meet the Department of Health's current policy objectives.

1.2 This report covers:

- background to the modelling technique (Part One);
- methodology (Part Two);
- results (Part Three); and
- discussion (Part Four).

Scope

1.3 We focused on the care during admission for potentially delivering the baby (intrapartum care). In other areas of the maternity pathway there is less certainty about how resources are assigned to activity and, therefore, modelling work would be less instructive.

1.4 We did not attach costs to the different resource uses. However, as raised in the discussion, we think this would be a good way to develop our work if it were to be applied elsewhere in the NHS. Indeed this work was not designed to conclude on how best to organise maternity services. This would require a full assessment of the various aspects of cost-effectiveness.

1 Comptroller and Auditor General, *Maternity services in England*, Session 2013-14, HC 794, National Audit Office, November 2013.

Selecting the NHS trust

1.5 We identified hospital trusts that offered a full range of maternity care settings, so that we could investigate the effect of having both midwifery-led and obstetric units. The selected trust was chosen as it could give us the detailed data required to create the model. The trust's maternity services consisted of:

- an urban hospital site with both an obstetric unit (OU) and an alongside midwifery unit (AMU); and
- a separate freestanding midwifery unit (FMU).[2]

1.6 The trust's maternity services delivered some 6,000 babies in 2012-13; less than 1 per cent of these were home births. We also included within the model the approximately 1,700 non-delivery episodes (where a woman was admitted to the hospital but did not give birth). These episodes created demands on the same resources as those used for delivering babies. The trust's key maternity service resources are set out in **Figure 1**.

Figure 1
Key maternity service resources

Resource	Capacity or number
Beds	32 in postnatal 15 in labour ward 10 in antenatal 5 in alongside midwifery unit 3 in freestanding midwifery unit
Midwives covering labour ward and theatre	11 (including 2 in triage) covering labour ward and theatre 4 covering the alongside midwifery unit during the day, 3 at night
Theatres	2: one 24/7 theatre; and one working 8 am to 1 pm

2 AMUs and FMUs are both types of midwifery-led units, with the former located on the same site as an obstetric unit.

Scenarios

1.7 We used the current level of care provision as the baseline scenario. We then compared this to a number of scenarios that altered one or more aspects of service provision, or demand, to demonstrate changes to the resource requirements. The scenarios are discussed in greater detail in Part Three of this report.

1.8 The assumptions used for changes in demand were based on the forecasts of the regional commissioning support service.[3] Specifically, the scenarios were:

- increased complexity (higher caesarean section rates);
- one-to-one midwifery care;
- increased number of deliveries;
- increased use of midwifery-led units;
- more home births;
- additional theatre use; and
- reduced length of stay.

Modelling method

1.9 We used a 'discrete event simulation' (DES) to model the provision of maternity services by replicating the current care pathway.[4] We based this on clinical guidance and consultation with staff at the trust. We used DES as we were modelling a complex environment with different patient profiles and resource constraints. DES also allows for modelling processes where there is variation in when events happen or how long they take. Furthermore, it should be easy to make changes in the model and observe the differences in the outputs.

1.10 DES models allow a real-time calendar to be used in modelling care scenarios. They also allow us to integrate resource availability into the modelled pathways. This lets us identify resource issues that the health system may face. In the case of maternity care, the resource constraints include the availability of beds, theatres and midwives.

3 These are commonly referred to as commissioning support units (CSUs). They are NHS-supplied business intelligence providers comprising data management and integration centres. The centres provide data validation, integration and storage to cleanse, validate and link national and local datasets. The datasets give meaningful conformed data that is then available for further analysis.
4 In this form of simulation, a system is modelled as a discrete sequence of events in time.

Part Two

Methodology

Model description

2.1 We developed a 'discrete event simulation' (DES) model to evaluate the resource implications for changes in maternity care provision and demand, using Simul8 software. When modelling the system, we made certain assumptions and simplifications to reduce complexity and account for any gaps in the evidence (**Figure 2**). We discussed the reasonableness of the assumptions and simplifications with healthcare and modelling experts.

Figure 2
Assumptions for creating pathways

Issue	Likely explanation	Assumption
We do not know if the same woman was discharged without giving birth and readmitted later when she eventually gave birth.	The data we had was anonymised.	We treat birth and non-birth as two different cases (and therefore the model can count the same woman more than once).
Some mothers had very high numbers of episodes within the same admission.	Some were likely due to extra coding while others were due to a small number of atypically complex cases.	We removed cases with six or more episodes, which accounted for only 2.4 per cent of mothers.
Some wards had very low usage (for example one episode at the radiology day unit).	Incorrect coding or very atypical cases.	We removed non-maternity wards with low usage.
The antenatal and postnatal rooms were used interchangeably.	During busy periods, to ensure all mothers have a bed these resources are pooled.	The model considered these rooms as a single ward, pooling both resources and the demand for these resources.
There were a wide range of methods of delivery.	Differing patient characteristics and case-mix require different clinical input.	We removed rare methods, for example only 0.2 per cent of the mothers had 'breech' as final method of delivery. We also merged some other methods that were 'clinically' similar, for example we grouped ventouse and forceps together.

Maternity pathways

2.2 We developed the care pathway model based on the limited existing literature and discussions with maternity specialists. We also visited the trust and reviewed a sample of patient notes to gain further insight into how the pathways are operated. The trust has a triage system to assess women on admission, which is staffed by two midwives. As stated previously, the trust has an OU, AMU and FMU. We modelled five different pathways (**Figure 3**). We compared the proportion of cases within each pathway for the trust, as given in Figure 3, to summarised Hospital Episode Statistics data covering other providers, and found that it is typical for an NHS trust.

Defining the pathways

2.3 After identifying these pathways, we defined the possible transfers from one facility (e.g. unit, ward, theatre) to another and the corresponding likelihoods for women following each separate path. We removed very rare transfers. We discussed these changes with the staff from the trust, who also suggested that some other transfers were very unlikely.[5] We removed such transfers from our model pathways, as they are not a clinically necessary part of the process of giving birth. We validated the pathways for each method of delivery with the help of trust staff. Two examples of the pathways that we created are presented in **Figure 4** on pages 10 and 11.

Figure 3
Model pathways

Pathway	Proportion of cases (%)
Spontaneous vertex (vaginal birth of a head down baby)	53
Emergency caesarean section	13
Elective caesarean section	6
Instrumental births (ventouse and forceps)	6
No birth (women that are admitted but do not give birth), for example, ‘false alarms’	23

5 For example, in the pathway for the spontaneous vertex, 4.1 per cent of women in the antenatal room were transferred to the AMU. The trust explained that this could happen only when it did not have available beds, and not for ‘clinical’ reasons.

Figure 4
Example pathways

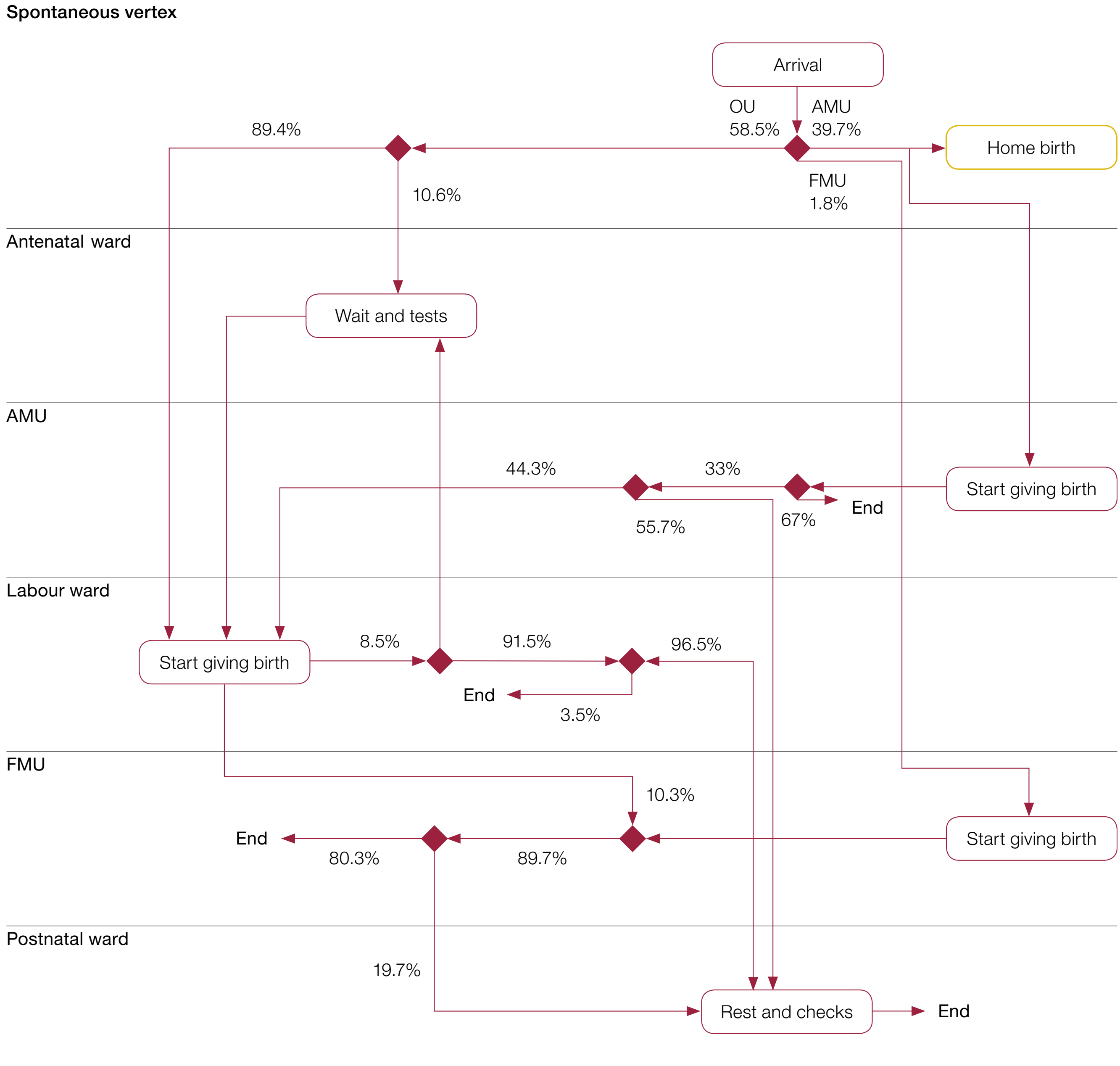
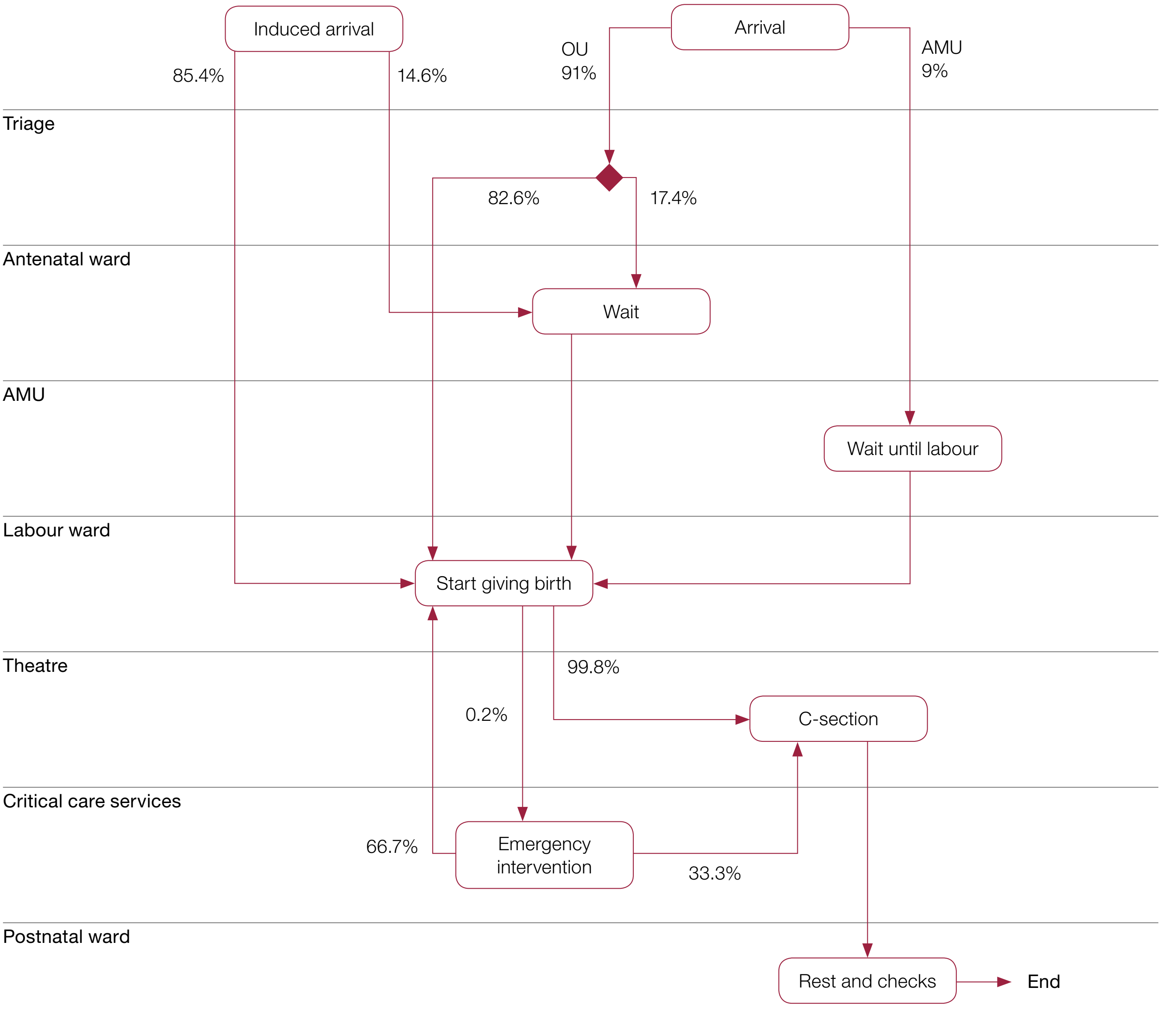


Figure 4 *continued*
Example pathways

Emergency caesarean sections



Data

2.4 We obtained information on women's progress through the maternity services from: the Patient Administration System (PAS); Evolution (the IT system used by the midwives); and the separate theatre dataset. From each source, we extracted data for 2011-12 and 2012-13. We undertook a range of data-cleaning exercises including:

- comparing the time stamps on the various data to evaluate which were the most realistic; and
- removing outliers that were likely to result from a coding error (for example, negative lengths of stay).

2.5 To ascertain the maximum available resources (beds, theatres, midwives), we used the trust's response to our survey of maternity services. We determined the 'resource rules' (availability throughout the day and when, during women's labour, that resources are required and released) by interviewing trust staff and considering National Institute for Health and Care Excellence (NICE) guidance.

Arrivals

Assumptions

2.6 We checked arrivals data for seasonality using the Chi-squared test. Although there were some associations between arrivals and month of the year or day of the week, we did not identify any time factor to be consistently significant across both years. So, to avoid unnecessary complication, we decided not to include seasonality either by month or day of the week.

2.7 To examine the admissions by time of day we removed elective caesareans and inductions as these women are scheduled arrive at 8 am. We still identified a spike in admissions at 8 am for the remaining cases. However, after discussing this with hospital staff we identified this was due to the trust's data system using an 8 am default time.

Defining arrival distribution

2.8 The next step was to define the arrivals in our model. For the elective caesareans we divided the total number of cases by the number of weekdays within a year and set two scheduled arrivals per day at exactly 8 am, except for weekends. Inductions can be scheduled during weekends as well so, in this case, we divided the total number of induced cases by 365 days. This calculation suggested that we should schedule three inductions every day.

2.9 For the other pathways, we tested what distribution best fitted the data. A Poisson distribution could describe the number of mothers arriving each day for all the paths, with an exponential distribution describing the time between each arrival. As none of the pathways demonstrated consistent seasonality, we grouped them together and used a single entry point for all mothers in our model. Our model then assigned mothers to the specific pathways according to the actual percentages seen in previous data. Specifically, mothers enter the model (at the entry point) with time gaps derived from an exponential distribution with λ (rate parameter) = 1.47 hours (equivalent to Poisson distribution with λ = 16.3 mothers per day).

Women's duration in different settings

Data cleaning

2.10 We removed outliers in length of stay after agreeing with trust staff that the outliers were incorrect values. Moreover, we included only the duration of any operation because this is the time the woman occupies the theatre bed. After this she spends some 45 minutes in the recovery room and is then transferred either back to the labour ward or the postnatal room.

2.11 We also considered 'hidden queuing times', especially for the labour ward. In other words, sometimes the woman cannot be transferred from the labour ward to the postnatal room because there are no available beds. In such instances the woman will wait in the labour ward longer than is clinically necessary. We were not readily able to ascertain how often this happens and, therefore, the difference between the two timestamps we have (entering the postnatal room minus entering the labour ward) includes both the actual duration of labour and this extra waiting time. To investigate how to correct these hidden queuing times, we examined two different approaches:

- Use the 'stage 3' of labour plus a fixed amount of time (trust staff suggested two hours) as the end time of the labour episode, instead of the timestamp that the mother enters the postnatal room. In that case, we could assume that the rest of the time, until the transfer, was a delay due to lack of beds. We rejected this solution, however, because we found that the stages of labour did not agree with the timestamps we had for the transfers.

- Identify the least busy month for the trust and estimate the duration of labour using that period. This is based on the assumption that there is no lack of capacity then, as it is not a busy period. To test this solution we plotted the time spent on the labour ward for each month. However, we did not observe any difference between months in terms of how busy the maternity service is. In addition, the hospital also confirmed that there are no noticeably less busy months. So we rejected this approach as well.

2.12 Consequently, we used the initial data, knowing that in some cases they may include some extra time. This means that the duration for women requiring a midwife in our model is likely to be an upper estimate. In other words, the clinical time (duration of the process of giving birth) a woman spends in labour can be less than the actual time included in our model. We recognise this as a limitation of the model and raise it, where appropriate, in relation to relevant results in the following section.

Case mix

2.13 After we finalised the data for estimating the length of stay in each facility, we moved on to identify what affects this duration. The factors we examined were: age of the woman, parity (whether she had a previous child or not), the facility she came from, whether she gave birth in the current facility or not, and the next facility or setting into which she was transferred. Using Analysis of Variance (ANOVA) tests to investigate significance, we decided to include the following factors (where applicable):

- where the woman is transferred afterwards (for example, the length of stay in the AMU is different if she is transferred to the labour ward compared with being discharged); and
- whether or not the woman was induced.

Distributions

2.14 We did not have data for home birth durations but the trust estimated that they last approximately six hours. Therefore, we used an Erlang distribution centred around six hours to include some variation. The remaining distributions were fitted using the software 'Stat-fit' (an add-in of Simul8). Initially we used the distribution that best fitted our data series, and when unable to fit a distribution we used historical data.

Resources

2.15 A summary of the resources in the trust is given in Figure 1. These resources – and how they are used in the trust – are described in more detail below.

Midwives

2.16 The trust operates with some flexibility in its staffing. A set number of midwives on shift are responsible for the women in labour and theatre, and respond to whoever needs care (depending on which midwife is available). When all these midwives are busy and there is a need for more staff, a midwife from the alongside midwifery unit (located on the same site as the obstetric unit) may be called. Midwives are not currently a strict resource constraint as they can take care of more than one woman at the same time, depending on how busy the hospital is. Consequently, we considered them as an 'infinite resource'. That is, they are not a constraint in the model in describing the current operation; but they are applied as a constraint when testing the one-to-one care objective. However, we do present the proportion of time that there are specific numbers of women in each of the wards/units, which can be used to inform local decisions on how many midwives would be required.

2.17 Some midwives are attached to specific facilities (antenatal, postnatal, AMU). Others work across more than one facility during the same shift (for example, labour ward and theatre). The only scenario for which we applied midwives as a constraint was relating to one-to-one care during established labour. For this scenario, we only considered midwives that worked in the labour ward (who also cover the theatre) and the AMU.

2.18 We did not differentiate the midwives by seniority. Finally, we assumed that there is no travelling time as the model considers that the midwives work across wards that are located on the same site.

Beds

2.19 The beds of the antenatal and postnatal room can be used interchangeably when there is no spare capacity. Moreover, we did not include any beds dedicated to inductions. However, the total number of beds can be used by all women whether they are induced or not.

2.20 We modelled the two available theatres as two separate resources, because they have different working hours. One is available all day every day, and the other is dedicated to elective caesareans from 8 am to 1 pm.

Approaches

2.21 We investigated two approaches to using resources. In the first, we included within our model a separate queue before every facility, while in the second we did not. This has a direct impact on using resources. In the first case, the woman leaves a facility (and releases the resources used) even if there are no available resources to be transferred to the next one. While in our second approach, the woman leaves a facility only when she can be accepted in the next one. In this case, even if the care in the current room has finished, she waits there (keeping the resource) until she can go on to the next facility. We decided to follow the second approach as this is more operationally realistic, based on our observations.

Model verification and sensitivity analysis

Warm-up period

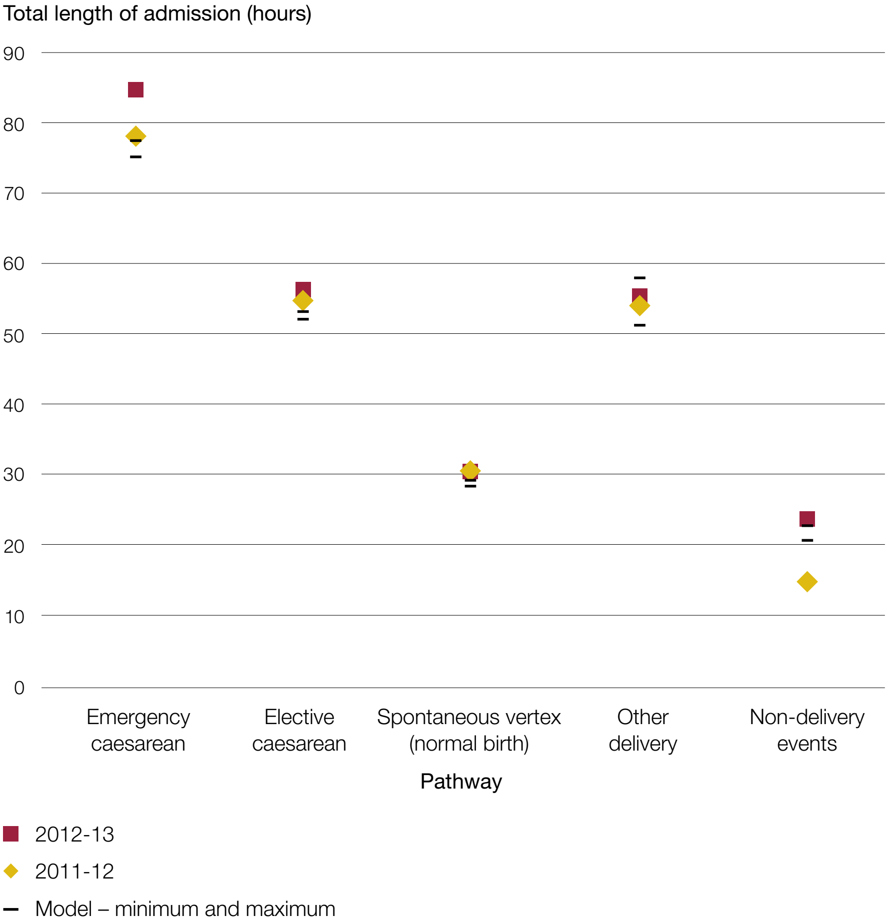
2.22 When testing alternative scenarios it is important to collect results using a consistent methodology. For the current state and the various scenarios the analysis should not start with an empty system. That is, the starting point for analysis should be when there are a realistic number of women in – or waiting to be in – the services. As a result, a warm-up period should be set to ensure cases are routed to the pathways as they would be in real life (with finite resource capacities). The key indicators we monitored were the total length of admission (the difference between admission time and discharge) for each pathway and overall. We used the maximum period across two different methods (Welch and Robinson) to reach a steady state to determine the warm-up period, which was eight days.

Validating model data

2.23 To check that the 'cleaned' data used in the model still represented the actual situation, we compared the medians (to reduce the effect of outliers) of the original and cleaned data. After running the model five times with different streams of random numbers we found the maximum and the minimum median, for each path. We cannot estimate a confidence interval using the median, so we used the medians of both years' real data, to observe how much variation we might have (**Figure 5**).

2.24 As we see, the model's outputs are very close to the real-life value. In most cases, the model outputs are slightly lower. This reflects how we cleaned the data, with unrealistically high values easier to identify (and therefore exclude) as errors compared to low values.

Figure 5
Validity of the model data



Expert panel

2.25 During the development phase, we presented the model to a number of modelling and clinical specialists for comment. This process enhanced the model's structure and we repeated it at each stage of the model's development. We acknowledge their input, for which we are grateful, at the end of this report.

Replicating the scenarios and sensitivity analysis

2.26 Simul8's trial calculator recommended using fewer than ten runs (to have 95 per cent confidence for the main key performance indicators, such as average length of admission). To increase the robustness of our results but keeping, at the same time, the running time of the model very short, we selected a larger number (14 runs).

2.27 We tested the robustness of the model's results when changing some of the inputs, by analysing the effect of different assumptions, such as: run lengths (1, 3, 6, and 12 months); pooling the antenatal and postnatal ward beds; and using one distribution family for the same facilities across all pathways and cases. This provided reassurance that the model was not overly sensitive to these assumptions.

Part Three

Results

3.1 For each scenario we set out: the context; how we tested it, including the assumptions we made; and the results.

Scenario one: Complexity

Description

3.2 The regional commissioning support service predicted that there will be more complex cases up to 2015-16. More precisely, it presents a scenario with caesarean sections (elective and emergency) increasing by around 10 per cent. This will result in a corresponding decrease in normal births.

How we tested it

3.3 We increased the number of women arriving in the pathways for emergency and elective caesareans by the forecast percentages (9.2 per cent and 9.8 per cent, respectively). In the case of elective caesareans we put one more arrival every week (for simplicity the extra arrival is not fixed at 8 am). To keep the same total volume of women, we assumed that all the extra caesarean cases came from the spontaneous vertex pathway.

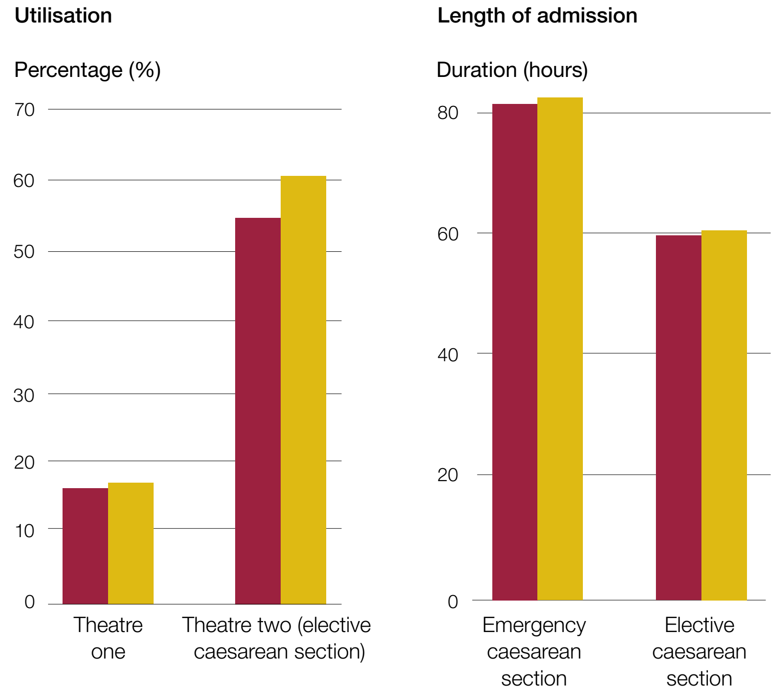
Results

3.4 As we expected, the model suggests increased use of both theatres (**Figure 6** overleaf). Also there is a 3 per cent increase in use of the postnatal ward. However, the finding that lengths of stay did not change notably suggests that the trust seems to be generally able to manage with this increased complexity. For both pathways there is a slight increase in lengths of stay because of an increase in the waiting time either to be admitted or to be transferred.

3.5 If the lack of capacity was significant, a queue of women would be created in the model, leading to a substantial increase in the total length of stay. For emergency caesareans the increase in length of stay was very small because these cases take priority over the other pathways when they demand a theatre bed. In addition, the risk that this priorisation (for emergency caesareans) adversely affects the length of stay for other pathways did not materialise, with slight decreases across these other pathways.

Figure 6

Increased complexity: resource use



	Theatre one	Theatre two (elective caesarean section)	Emergency caesarean section	Elective caesarean section
Initial (%)	16.5	55.4	81.1	59.3
Scenario (%)	17.8	61.4	81.3	59.8

Scenario two: One-to-one care

Description

3.6 In this scenario we test the one-to-one care objective. This means that throughout the process of giving birth, whenever a woman needs care the same 'named' midwife is responsible for providing it. In reality, this is inherently difficult to guarantee as there are more women than midwives on the labour ward. So, when a mother needs a midwife, the one that is least busy will be most likely to respond.

How we tested it

3.7 To test this we had to rely on one assumption: when a woman is in labour, theatre or AMU, for the whole period she 'takes up' a midwife. In reality, there are times that a woman in labour does not need a midwife. But, according to the objective, the midwife needs to be available. Moreover, it is not easy to predict when a woman will need the midwife. So, the midwife must be on standby for the whole time the woman she is assigned to is in labour.

3.8 In testing this scenario we kept the following points in mind:

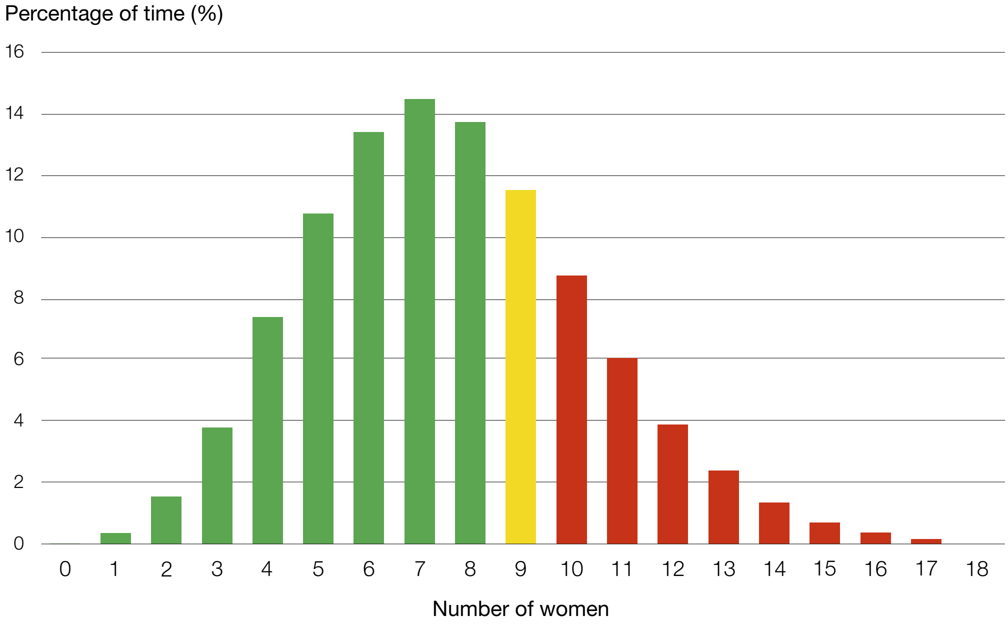
- The durations for the labour ward include 'hidden delays' (paragraph 2.10). In other words, the durations we have for labour episodes are always biased upwards. So, in our model a midwife appears to be needed for a longer time than in reality. Moreover, in the AMU a woman can receive care both before and after established labour as well, during which a midwife is not necessarily needed under the guarantee of one-to-one care during established labour. But in our model we did not have data to separate these times. As a result a midwife appears to be required for the whole time.
- Conversely, for some births, two midwives may be required, making the one-to-one care objective more difficult to achieve than the model estimates.

3.9 In this scenario we used 'infinite capacities' to examine for what percentage of time the number of women in the labour ward and AMU is higher than the number of available midwives. During these periods, at least one midwife is responsible for more than one woman at the same time. As a result, if all of the women need care simultaneously, another midwife (not necessarily the assigned midwife) will respond. We preferred this way of testing the objective, because it reflects actual practice more closely. Although the trust tries to provide one-to-one care whenever possible, it will admit a woman if there is not a midwife to take exclusive care of her.

Results

3.10 We observe that for about 23.5 per cent of the time (the red bars in **Figure 7** overleaf) there are more women in labour and theatre than available midwives. An additional three midwives would allow for one-to-one care in all but 5 per cent of the time.

Figure 7
Women in labour ward



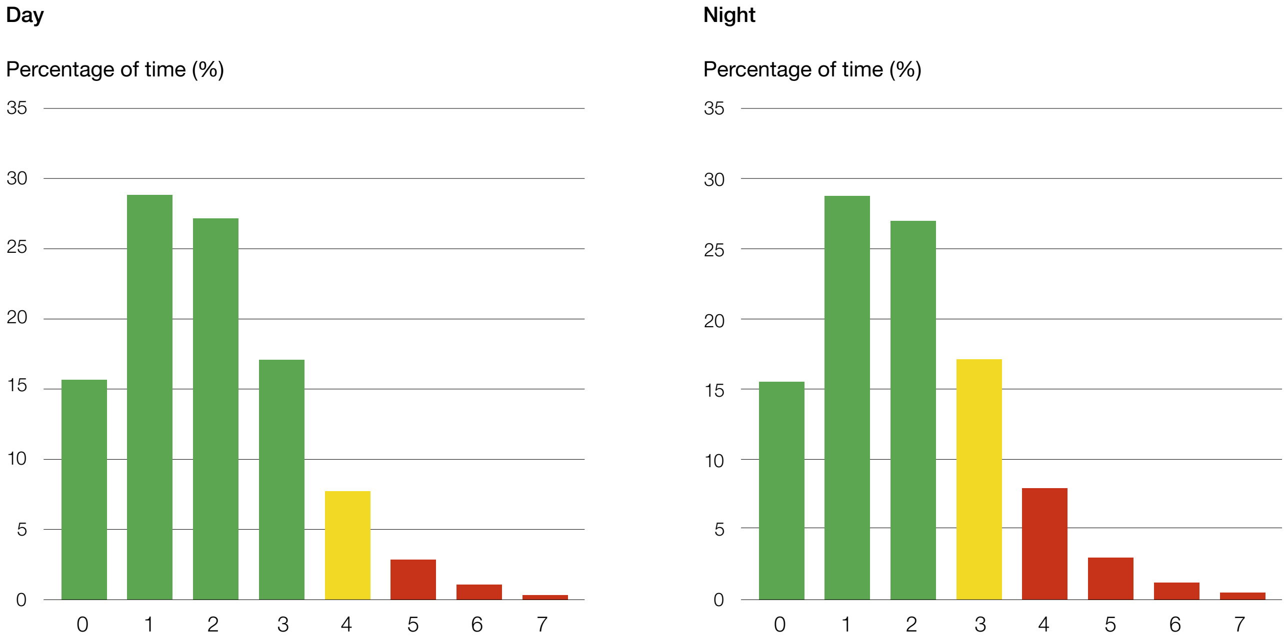
Note

1 Amber bar shows the current number of midwives at these times.

3.11 Even if this includes the limitation of the 'hidden delays', the objective of one-to-one care cannot be achieved at all times. This is because in many cases a midwife must be assigned to two women at the same time. And if these women require the presence of the midwife simultaneously, the objective cannot be met.

3.12 On the other hand, the results on the AMU show a greater ability to provide one-to-one care (**Figure 8**). In this case, to be able to comment, we had to separate the total time into hours of day and night, because there are a different number of midwives available. In particular, during the day the percentage of time that the number of women exceeds the number of midwives is very small (4 per cent). During the night the percentage is greater (11.8 per cent) because there is one fewer midwife. If we also consider that durations within the AMU included time (for post-labour care) where the one-to-one objective is not relevant, the AMU is probably well placed to meet the objective nearly all of the time.

Figure 8
Women in alongside midwifery unit



Note
1 Amber bar shows the current number of midwives at these times.

Scenario three: Increased numbers of women

Description

3.13 In this scenario we were interested in testing whether the system can cope with an increase in demand. This can be caused, for example, by closing another local hospital. Because of the small total length of stay we were not interested in observing the system in the transient state (during the increase in numbers). Instead, we wanted to test the steady state (when the full extent of the increase in demand is affecting the service).

How we tested it

3.14 To simulate the increase in demand, we increased the number of all arrivals by 10 per cent (as forecast by the regional commissioning support service). For the elective caesareans we put one more arrival per week. For inductions, we put one more arrival every four days. For simplicity the extra arrivals are not fixed at 8 am.

Results

3.15 We observe an increase in demand, resulting in an increase in use, as expected. The two theatres have the greatest proportional increase (**Figure 9**).

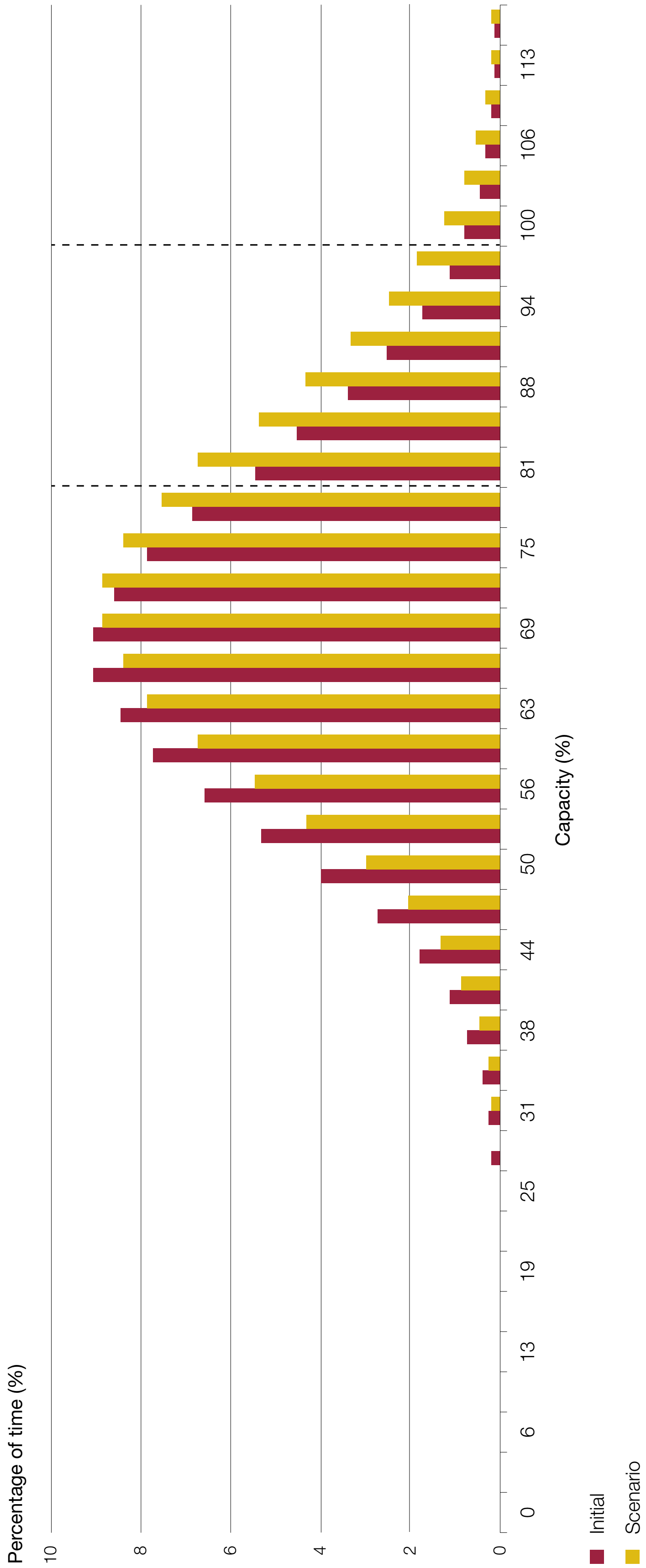
3.16 The facility that raises the most concerns is the postnatal ward because it was already at 69 per cent occupancy. With the increase in numbers, occupancy of postnatal beds goes up to 74 per cent. This is approaching the Trust's internal policy of no more than 80 per cent occupancy. In summary, the Trust does seem capable of absorbing the increased demand within its existing capacity (bed numbers). However, there might be potential capacity issues in the postnatal ward (**Figure 10**).

3.17 The increase in theatre use is also reflected in an increased total length of stay for elective caesareans (up by 1.6 per cent). Other lengths of stay remain consistent.

Figure 9
Increased numbers of women: resource utilisation rates

Resource	Initial (%)	Scenario (%)	Change (%)
Freestanding midwifery unit (FMU)	1.9	2.0	5.3
Alongside midwifery unit (AMU)	40.9	44.2	8.1
Antenatal	50.4	54.3	7.5
Labour	51.3	56.4	9.8
Theatre	16.5	18.3	11.0
Postnatal	68.8	74.5	8.2
Theatre (Elective)	55.4	62.6	13.0

Figure 10
Postnatal ward use



Note

1 The dashed line represents 80 per cent and 100 per cent demand for beds in the postnatal ward.

Scenario four: Increased use of the freestanding and alongside midwifery units

Description

3.18 This scenario represents the prediction that there will be increased used of the freestanding and alongside midwifery units. In this scenario we were interested in observing how this would affect the operation of the hospital as well as the midwifery-led units.

How we tested it

3.19 To test this we changed the number of arrivals of the AMU, FMU and hospital (triage) by increasing the women that go to the AMU by 11 per cent and to the FMU by 2 per cent. The corresponding number of women were removed from the obstetric unit pathway. We did not change the numbers of induced cases, because these are not conducted in the midwifery-led units.

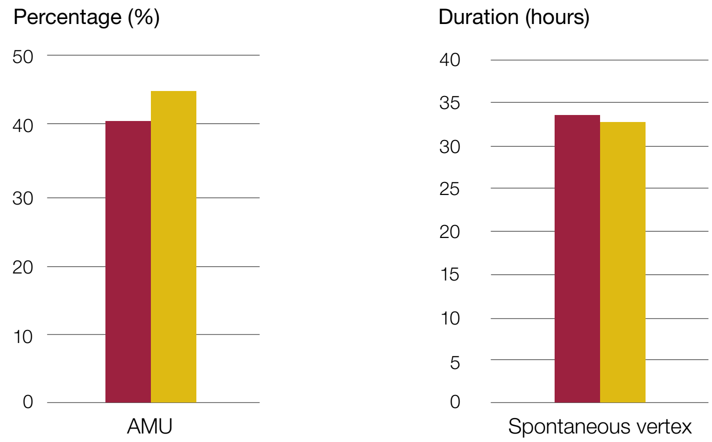
Results

3.20 As we expected the use rate of the AMU is increased by exactly 11 per cent. This happens because we increased the patients arriving at the AMU by that number and women cannot arrive as a transfer from another facility. The increased use of the FMU (which was easily able to absorb the additional demand for services as it currently runs at a low occupancy rate) and AMU also leads to a reduction in total length of stay (especially in spontaneous vertex) (**Figure 11**). This is because the data suggests women spend, on average, less time during their delivery admission in the midwifery-led units than in the hospital.

3.21 On the other hand this increase in use leads to a larger percentage of time that the AMU needs more capacity: from 1.2 per cent of the time initially to 2.6 per cent. Although, on average, the use of the obstetric unit's wards is reduced, the percentage of time these rooms need more beds does not change substantively. In other words, the increase in use of midwifery-led units reduces, on average, the occupancy of hospital beds but it does not substantially reduce the peaks in demand for hospital resources.

Figure 11

Use and length of stay at the alongside midwifery unit



	AMU	Spontaneous vertex
Initial (%)	40.9	33.5
Scenario (%)	45.3	32.8
Change (%)	10.8	-2.2

Scenario five: Home birth rate

Description

3.22 In this scenario we wanted to test how the home birth rate affects the system. More precisely, we want to see what would happen if the trust achieved the NCT's suggested 'critical mass' of 5 per cent.[6]

How we tested it

3.23 In this case for our initial results we included the current – almost low – number of home births (fewer than 50, representing < 1 per cent of the trust's birth). To test the difference we increased the home births to 5 per cent. We took these cases out from the arrivals at the AMU as this was deemed the most likely alternative place of birth for women who would consider a home birth (as the number using the freestanding midwifery unit is already very low).

6 The theory is that if a trust cannot achieve a 'critical mass' of home births, low local awareness coupled with limited midwife experience and resources may affect whether such services are offered consistently and could make them unsustainable.

Results

3.24 As in the previous scenario the most interesting results are again about:

- the use of the AMU (decreasing from 38.8 per cent to 37.3 per cent); and
- the total length of admission for spontaneous vertex (decreasing from 33.6 hours to 32.8 hours).

However, these results show that even a substantive increase in home births (to 5 per cent) is unlikely to have a significant effect on the operation of the trust's maternity services.

3.25 The other pathways are not affected because the difference in the number of home births is relatively low compared to the total number of births. So, relatively few resources are released by the increase in home births. This explains why we do not see a decrease in the 'blocked time'. In other words, the time that there is no capacity for a woman to go to the next ward, which would increase the total length of stay.

Scenario six: Two 24-hour theatres

Description

3.26 We wanted to test how the quality of service would improve if the second theatre (for elective caesareans) was available all the time instead of 8 am to 1 pm.

How we tested it

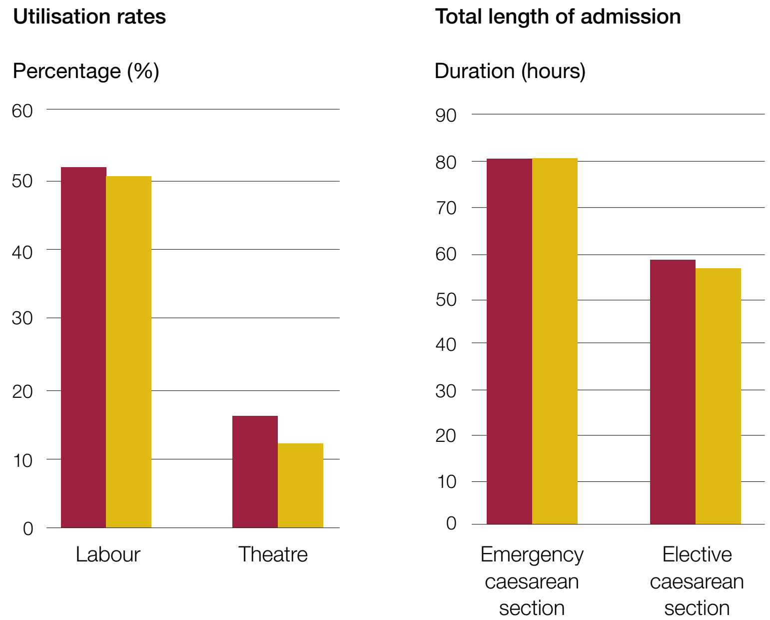
3.27 We increased the 24-hour, seven-day-a-week available theatre beds from one to two and took out the theatre dedicated to elective caesareans (instead these women are treated in one of the two 24-hour, seven-day-a-week theatre beds).

Results

3.28 Since we extend the hours that the second theatre is available without increasing demand for these services, we expected to have a reduction in utilisation rates of the theatres. The decrease in occupancy of the labour ward can be explained through this being the facility used before a mother is transferred to the theatre (**Figure 12**). With only one theatre, some women due to be transferred from labour to theatre would have to wait in the labour ward longer than they needed to. With two theatres, these cases are less likely.

3.29 In this scenario fewer women have to wait to go to theatre. This increases the quality of services with the decrease in the waiting time reflected also in the total length of admission, especially for elective caesareans. The difference in length of stay for emergency caesareans is small as they always have priority over other women requiring the theatre. And, furthermore, situations where there are two simultaneous cases are infrequent. From the graph below we see that two theatres are needed at the same time (for non-scheduled operations) in 1.2 per cent of the time while in nearly no circumstances are more than two required (**Figure 13** overleaf).

Figure 12
Theatre utilisation

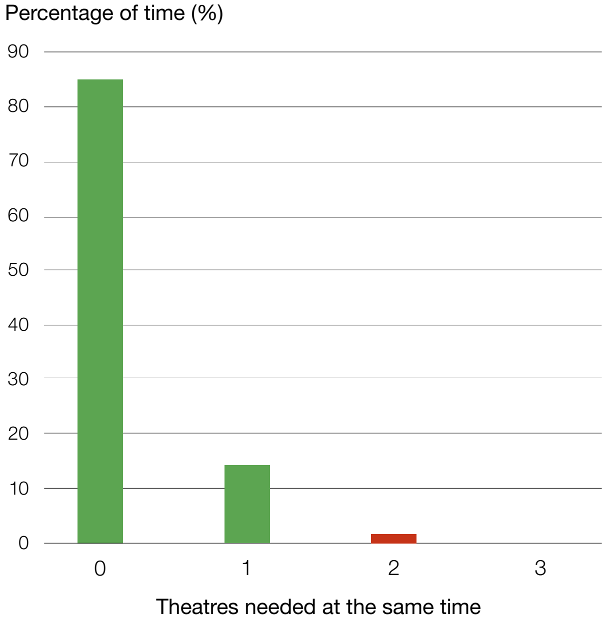


	Labour	Theatre	Emergency caesarean section	Elective caesarean section
Initial (%)	51.3	16.5	81.1	59.3
Scenario (%)	50.1	12.4	81.0	56.5
Change (%)	-2.3	-24.9	-0.1	-4.8

Figure 13

Simultaneous emergency cases requiring theatres

Theatres (apart from elective caesarean section)



Note

1 Bars shown in red indicate when current theatre capacity is exceeded.

Scenario seven: Reduced length of stay

Description

3.30 In this scenario we examined how using the resources would change, if the trust reduced women's length of stay. The regional commissioning support service predicts significant reductions to 2015-16 – a 45 per cent reduction in length of stay for spontaneous vertex, 35 per cent for instrumental births (forceps), 26.5 per cent for emergency caesareans and no change for elective caesareans. In this scenario we examined what would happen if the trust achieved half of these predictions across the different pathways.

How we tested it

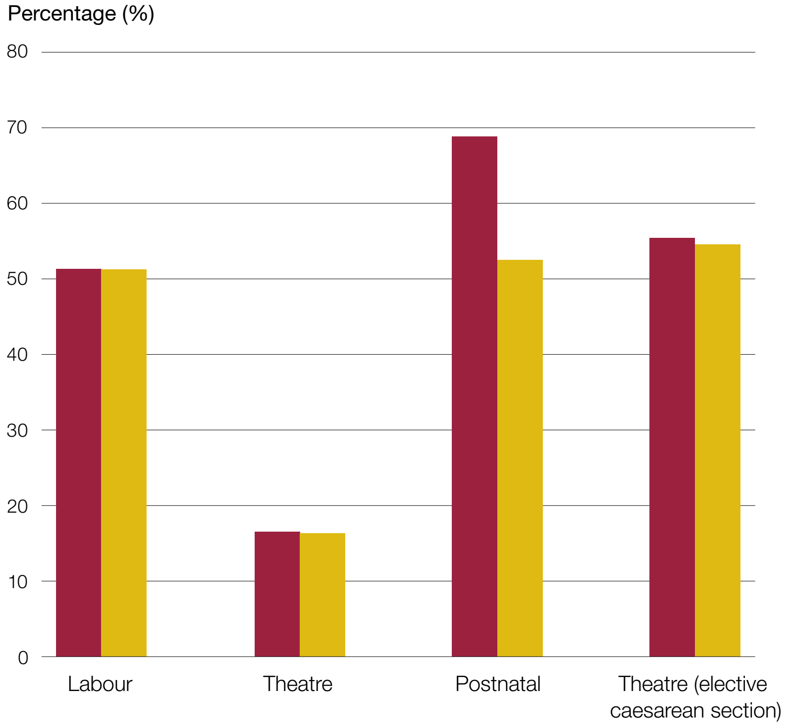
3.31 We assumed that the reduction in time can happen only in the postnatal ward. In all other facilities there can be no reduction, because medical procedures take place or lengths of stay are determined by a woman's progress through labour. So, we reduced the duration in the postnatal ward as much as it was needed to achieve the predicted decrease in total length of admission.

Results

3.32 We do not present the total length of stay for the pathways because they are purposefully set to the assumptions for this scenario, and instead present utilisation rates.

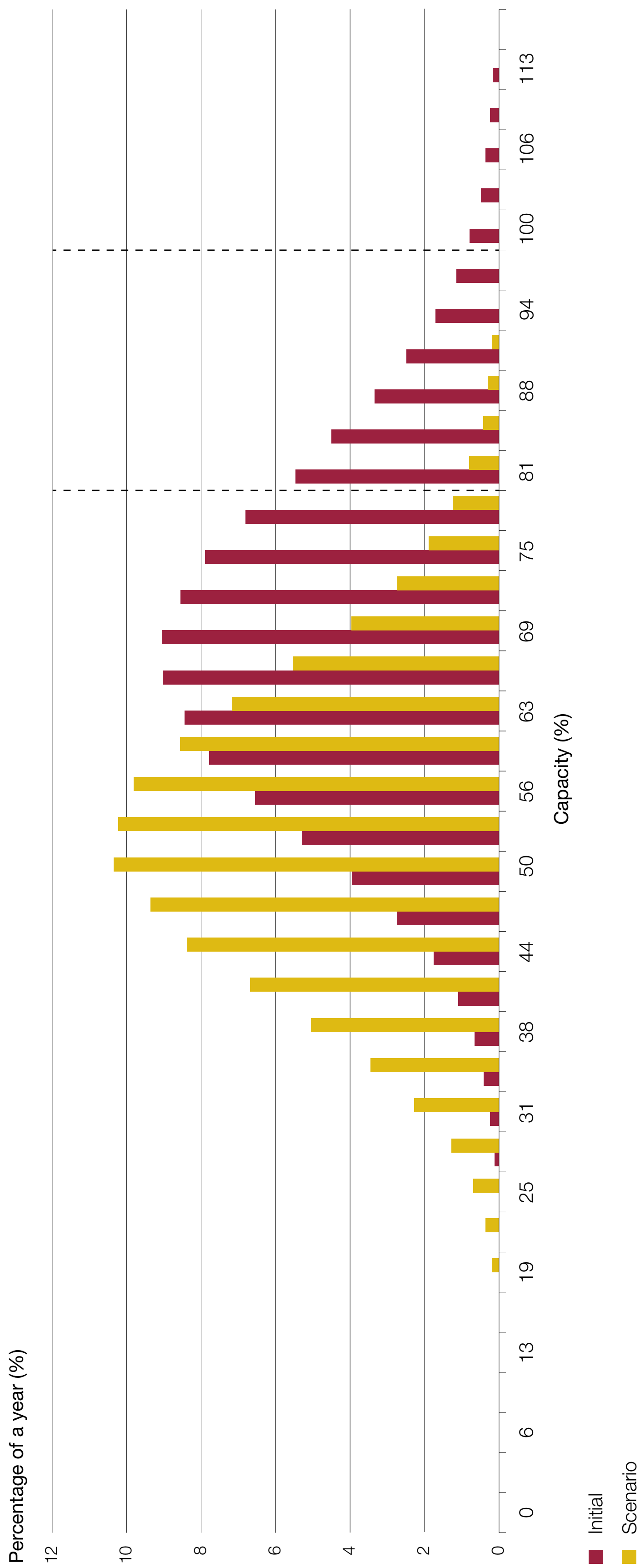
3.33 The greatest impact of reductions in length of stay is on the utilisation rates of the postnatal ward and, as a result, the facilities preceding the postnatal ward (albeit to a much lesser extent) (**Figure 14**). The postnatal ward is affected directly by the decreased time the mothers spend there. As a result, we observe a significant drop in the utilisation rate. In fact, we observe it is highly unlikely that the number of women requiring postnatal ward beds will outstrip the number of beds (**Figure 15** overleaf). The Trust could – in this scenario – consider reducing capacity.

Figure 14
Utilisation rates



	Labour	Theatre	Postnatal	Theatre (elective caesarean section)
Initial (%)	51.3	16.5	68.8	55.4
Scenario (%)	51.2	16.3	52.5	54.5
Change (%)	-0.2	-1.6	-23.7	-1.7

Figure 15
Bed demand in the postnatal ward



Note

1 The dashed lines represent 80 per cent and 100 per cent demand for beds in the postnatal ward.

Part Four

Discussion and conclusion

4.1 The modelling exercise, supporting our 2013 report *Maternity services in England*, took place between July and September 2013.[7] This report gives additional detail on the methodology and findings of the work. The findings were particularly relevant for the section of the report on 'managing local resources' which stated that:

- To provide an insight into the issues the NHS faces in managing resources at local level, we undertook some modelling work with a single trust. While the specific results are not generalisable to other providers, this work highlighted the challenges trusts face, given the fluctuations in activity, in providing a consistent service without having excess capacity. Key points to emerge were as follows:
  - Relatively large changes in activity (increases of 10 per cent in numbers of women and caesarean sections) could be handled within current resources, albeit with some additional pressure on certain beds (postnatal ward), and fully managed through a reportedly achievable decrease in lengths of stay.
  - In the labour ward, one-to-one care was achievable for around three-quarters of the time (broadly in line with the national average). Three extra midwives on this ward (an increase of a third) would be required to provide one-to-one care for 95 per cent of the time.
  - Although the emergency theatre was used for only 14 per cent of the time, there were rare occasions when an additional theatre was needed (1 per cent of the time).

4.2 The challenges highlighted by our modelling work are particularly acute for smaller maternity units – typically freestanding midwifery units – where the random variation in births is likely to have greater impact. It can be difficult for small units to ensure that staff are deployed effectively at all times. In response, some trusts are seeking to use resources flexibly, with staff providing antenatal and postnatal care in the community when there are fewer births. Other trusts rotate staff between different maternity units to ensure midwives are used effectively and gain sufficient experience.

7 Comptroller and Auditor General, *Maternity services in England*, Session 2013-14, HC 794, National Audit Office, November 2013.

4.3 This work requires detailed data and input from a range of trust staff. However, such techniques can give important insight into understanding risks and providing a cost-effective service that matches resources and likely demand. We recommend that the NHS considers using such approaches to plan and manage their services. However, to plan services there should be a wider assessment of cost-effectiveness rather than just resource use (as presented here). As such, further development of this approach should include attaching costs to the different resources.

Acknowledgements

The National Audit Office would like to thank the following experts:

- Dr Mike Allen
- Dr Tracey England
- Matthew Maguire
- Elaine Newell
- Professor Michael Pidd

National Audit Office

Design and Production by NAO Communications
DP Ref: 10259-002