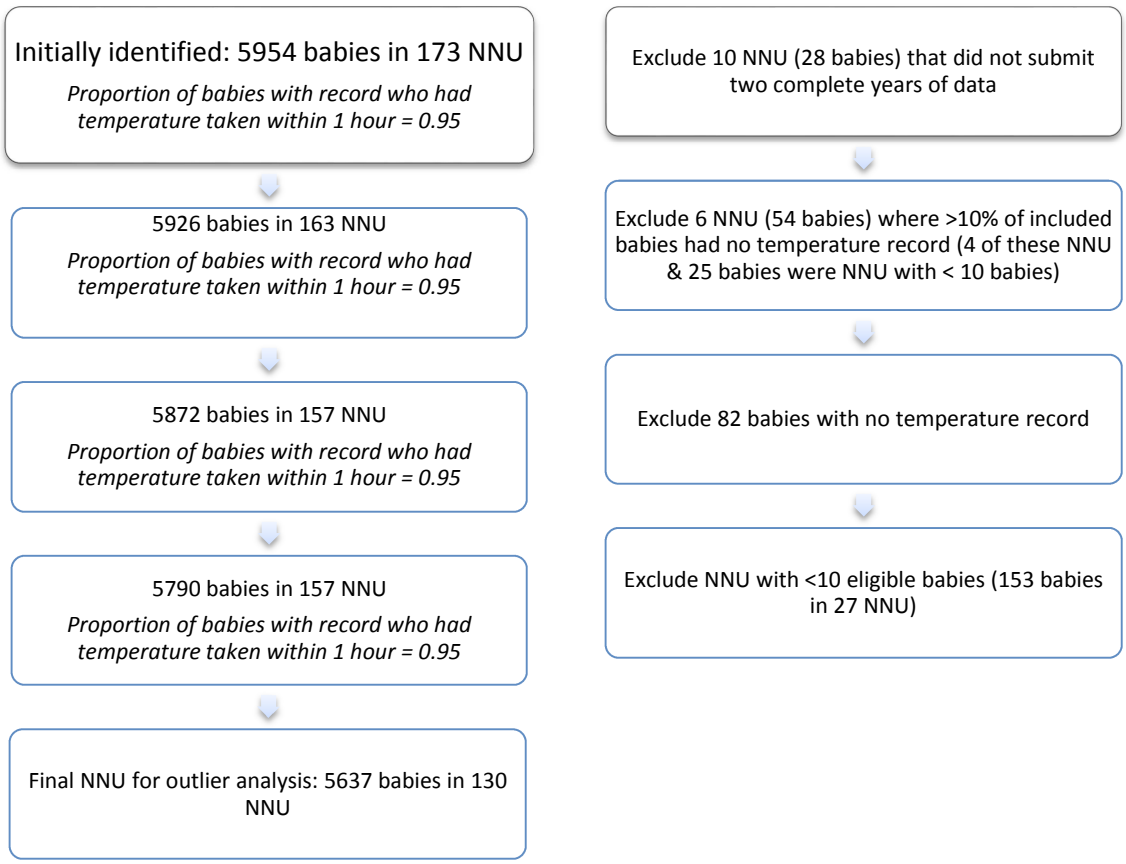
NNAP 2014 data full outlier analysis methodology and results

2014 Outlier Analysis for the NNAP audit measure on: Babies born before 29 weeks' gestation having temperature taken within one hour of birth. 2013 and 2014 data combined.

1. *Data set for outlier analysis:*
   This includes only babies born in hospital.

Figure 1: Neonatal Units (NNU) and babies included in the outlier analysis



This report is based on data submitted to the National Neonatal Research Database (NNRD) before the 18 June 2015. An initial analysis was done and reported to NNU using data submitted before 9 March 2015, after which NNU were permitted to change their data.

For the NNU with small numbers of babies included in this question, we have very little power to detect large deviations from what is usual in the population. In other words, the probability of detecting a large true deviation is very low when the number of babies is small. Applying the outlier analysis to NNU with small numbers of babies may lead to false complacency and/or inability to detect good performance. Therefore we based our analysis on combined data from 2013 and 2014, and we tested only the NNU with at least ten babies identified for inclusion in this audit question (see bottom right hand box in Figure 1).

Only NNU that had two complete years of data were included. There were ten NNU that did not submit two complete years (2013 and 2014) of data:

- Bronglais General Hospital
- Eastbourne District General*
- Harrogate District Hospital
- Hereford County Hospital
- Oxford University Hospitals, Horton Hospital
- St Richard's Hospital
- The Royal Free Hospital
- Torbay Hospital
- Withybush Hospital
- Ysbyty Gwynedd

*****Closed during 2013

Between 2013 and 2014, there was an improvement in the recording of temperature – with 2.1% of eligible babies having no record in 2013, and 1.1% having no record in 2014. In 2013, 95% of babies with temperature records had their temperature taken according to the NNAP standard, and in 2014 the rate was 96%.

The NNU with small numbers of babies, that were therefore not tested, are listed in Table 4 at the end of this document. There were 31 NNU with fewer than ten eligible babies – this includes the four small NNU with more than 10% missing temperature records.

The NNU with fewer than ten eligible babies had overall a lower proportion (91%) whose temperature was correctly taken on time, and this may be of concern - but even combining two years of data many NNU still have very small numbers of babies, and our tests of these small NNU on an individual basis lack power.

There were two NNU with at least ten eligible babies whose temperature-taking data is considered insufficient to participate in the outlier analysis, due to missing records, and these are summarised in Table 1. In the current report, the two NNU classified as non-participating would be included in the analysis if they were each to submit temperature records for one additional eligible baby with no recorded temperature in the NNRD.

Table 1 also shows the proportion of babies with temperature taken correctly under the worst case scenario that all missing records were due to babies not having their temperature correctly taken on time (“worst case”), and under the best case scenario that all missing records were due to babies having their temperature taken correctly on time (“best case”). We shall return to the non-participating NNU at the end of the outlier analysis of participating NNU.

Table 1: Non-participating NNU, i.e. those with > 10% of eligible babies having no recorded temperature

	Number eligible babies	Number with no recorded temperature	Proportion with recorded temperature	Number with temperature on time	Worst case	Best case
Princess Royal University Hospital	16	2	0.88	12	0.75	0.88
Wrexham Maelor Hospital	13	2	0.85	11	0.85	1.00

2. Outlier analysis

Methods:

The data set used for the outlier analysis is shown in Figure 1. The proportion of babies whose temperature was taken within one hour in each participating NNU was based on babies with temperature records only (where none of the participating NNU had more than ten percent of their temperature records missing).

We used a staged methodology to identify unusual performance, based on Ohlssen et al.[1] The first step in our staged methodology to identify potentially unusual NNU is an initial screen. In this step we used all the complete records in participating NNU to fit a model for appropriate temperature taking in the population (Figure 2). We used a random effects model, which allows for variability between NNU. We then used the first stage model's 5% prediction limit to flag NNU with low rates of appropriate temperature taking.

However, the first stage model is based on all the NNU, including those that may be unusual. The second stage of analysis comprises further testing of the NNU, by fitting a second model for temperature taking in the NNU that were not identified as unusual at the first stage (Figure 3). Using the second stage model, which allows smaller variation among NNU to be considered usual, we identify potentially unusual NNU using the standard limits: (alert) 2.5% and (alarm) 0.1%. The resulting funnel plot is useful for individual NNU to compare their performance to that of other units.

The final stage is to adjust for multiple testing. The final funnel plot is used by NNAP to identify possibly unusual NNU, allowing for the fact that 130 tests are made, and with each test the probability of incorrectly identifying a NNU as unusual is 2.5%, so that without adjustment the probability of incorrectly identifying a NNU as unusual is approximately 96%. To avoid identifying a large number of false positives, we control the expected proportion of false positives (this is known as the false discovery rate). We illustrate this using a funnel plot that shows the thresholds for false discovery rates 10% and 1% (Figure 4).

We note that the thresholds we have used are decided *a priori*, and are consistent with those used by national audit programmes, including past NNAP analyses, and in the literature on identification of unusual performance. Especially for NNU close to a threshold, the choice of

the cut-off may determine if the NNU is flagged or not. We emphasise that this process is useful to flag potentially unusual performance, but further non-statistical investigations of flagged NNU are necessary.

Results

NNU below 5% threshold: We identified four of the participating NNU below the 5% threshold at the initial screen. Figure 2 is a funnel plot showing the 5% threshold for the first stage model, and highlighting the four NNU flagged at this stage.

Two NNU were identified as below the 2.5% threshold for the second stage model – the accompanying funnel plot is shown in Figure 3.

Figure 2: Funnel plot for proportion of babies with temperature records who had temperature taken according to NNAP standards. The 5% control limit for the stage 1 model is shown, and NNU identified as possibly unusually low are highlighted.

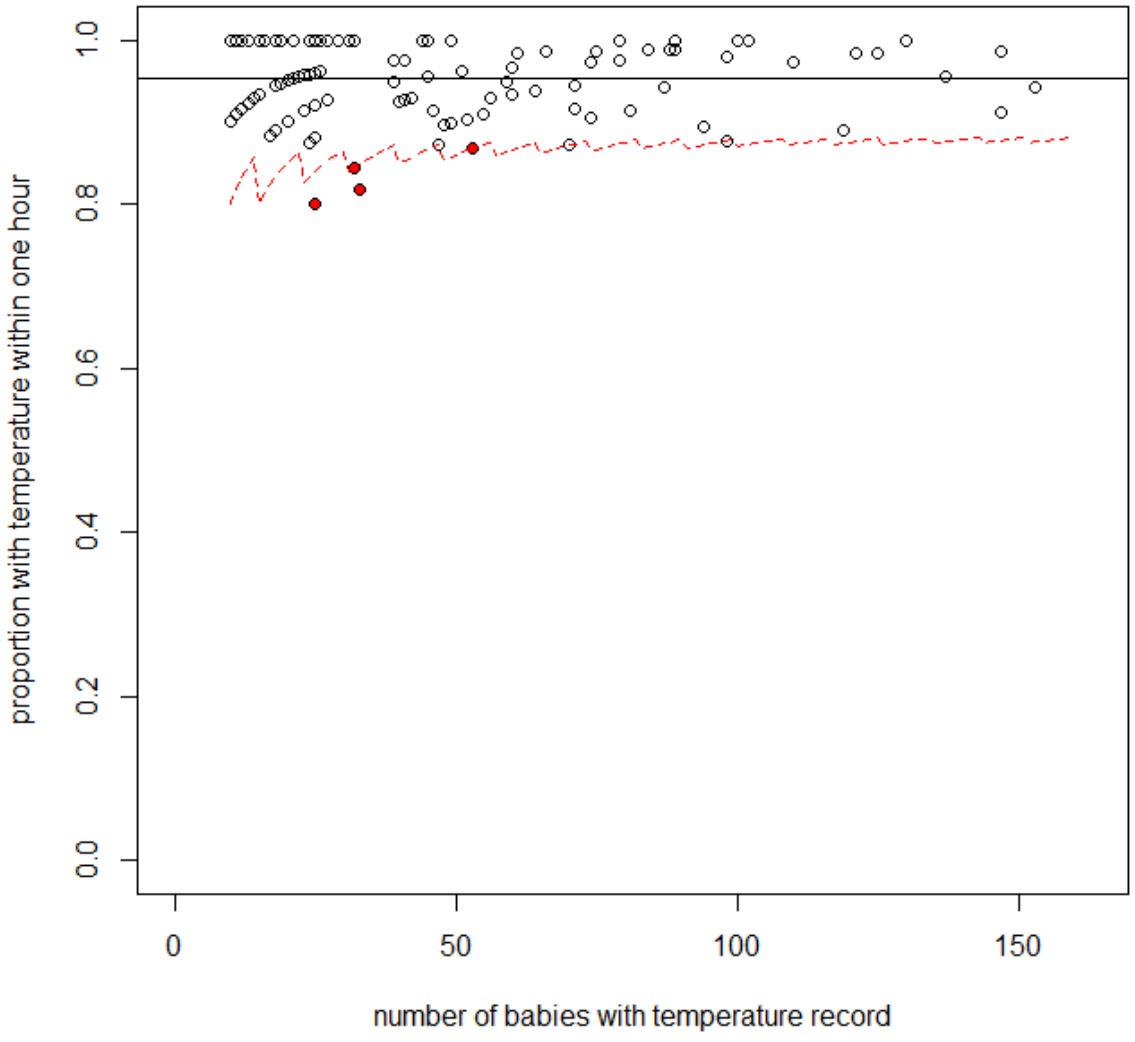
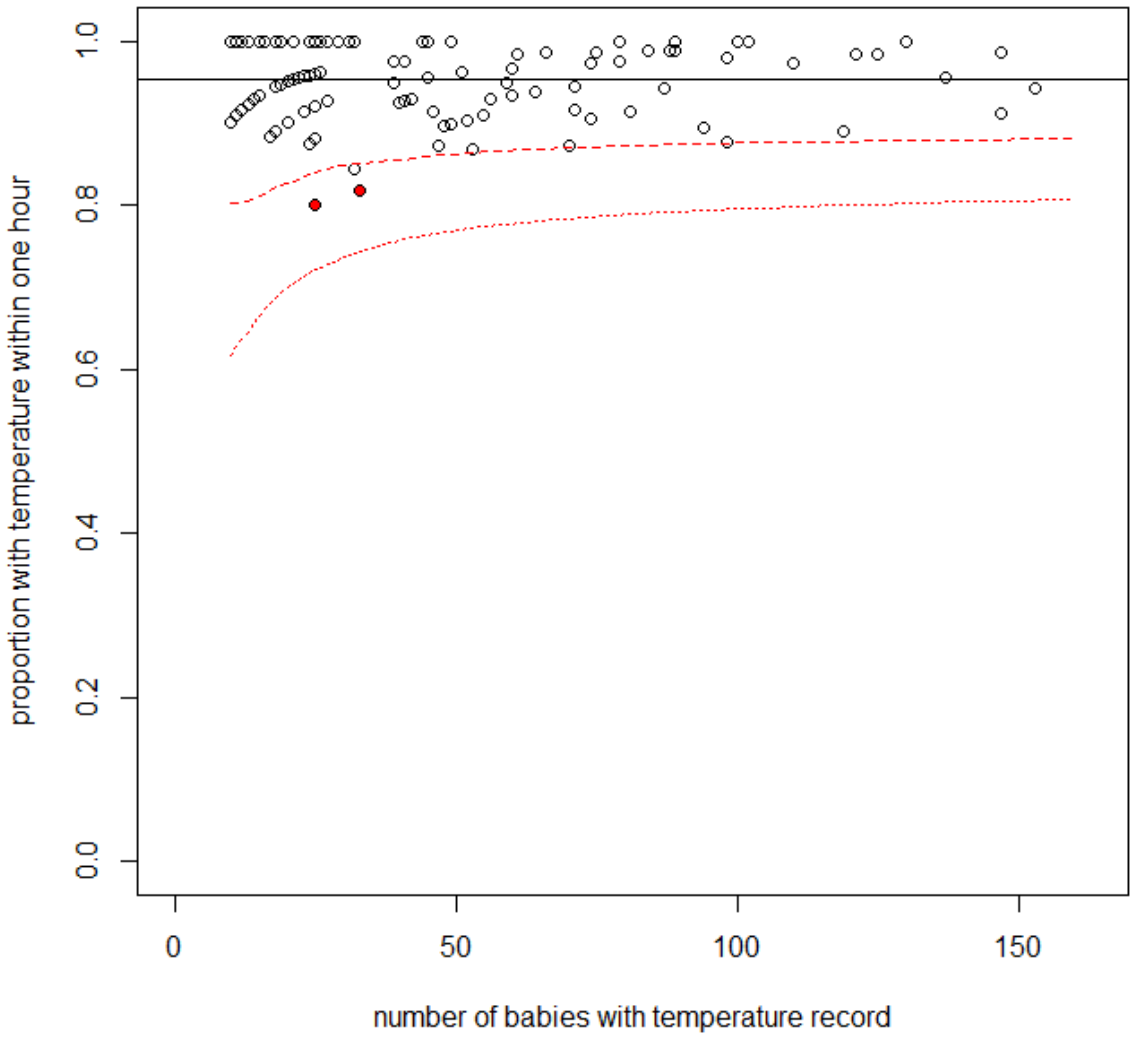


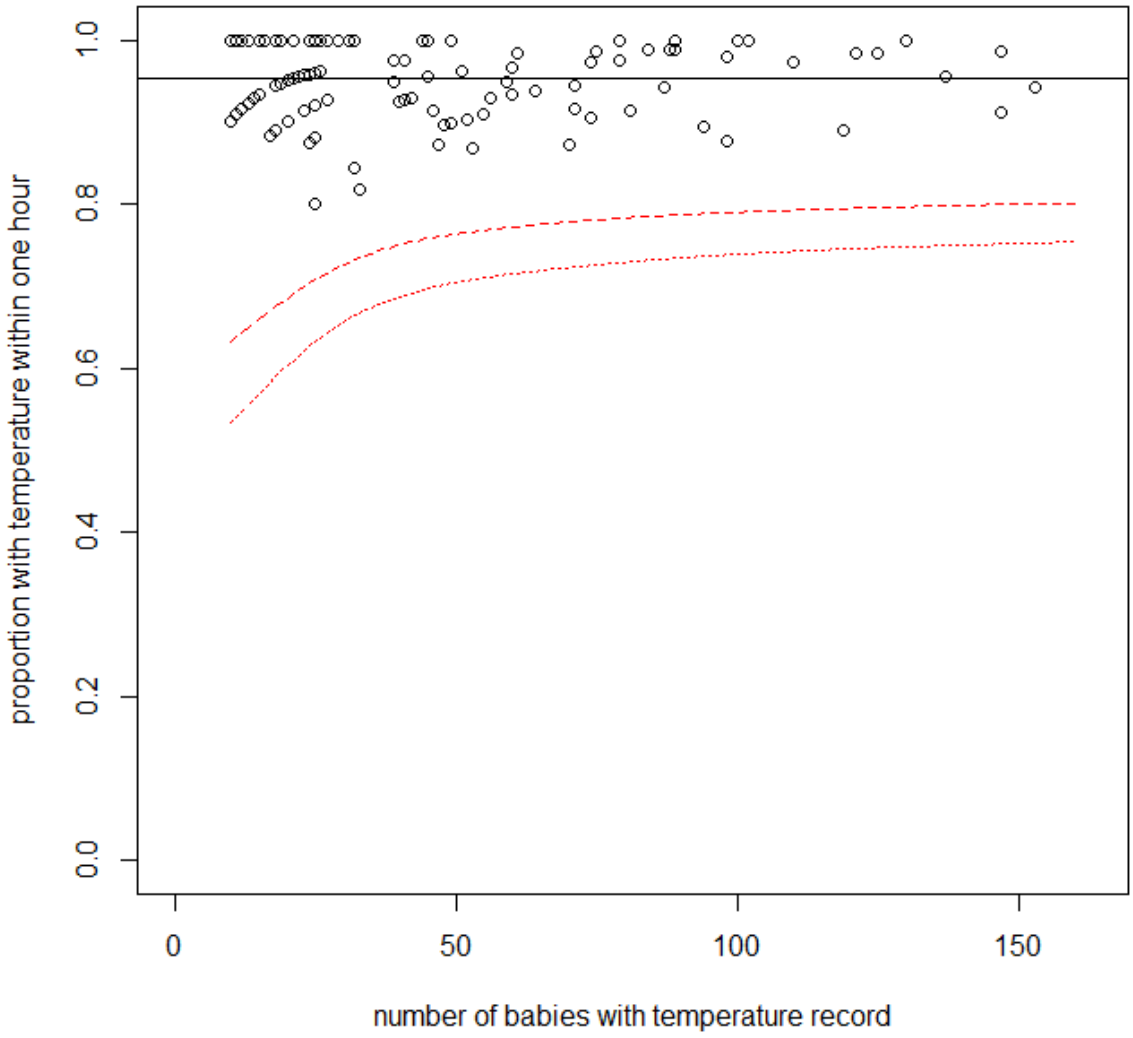
Figure 3: Funnel plot for proportion of babies with temperature records who had temperature taken according to NNAP standards. The 2.5% and 0.1% control limits from the stage 2 model are shown, and NNU identified as possibly unusually low for an individual test are highlighted.



Confirmation of 2014 low-outliers

On constraining the expected proportion of false positives to 10%, we found that no NNU would be identified as low outliers for this NNAP audit measure. This is illustrated in the funnel plot of Figure 4.

Figure 4: Funnel plot for proportion of babies with temperature taken within one hour of birth. The thresholds corresponding to the expected percentage of false positives 10% and 1% are shown.



Sensitivity analysis for missing values in participating NNU: We calculated the proportion with temperature taken on time in each NNU under the worst case scenario that all missing values were babies with temperatures not taken within one hour. We did not find any NNU would fall below the thresholds adjusted for multiple testing in the worst case scenario. This is not unexpected, as the proportion of missing records in the participating NNU is at most 10% and usually much less.

Sensitivity analysis for non-participating NNU: in the worst case where all the missing records were babies with temperatures not taken on time (shown in Table 1), neither of the two non-participating NNU would fall below the thresholds adjusted for multiple testing. One NNU would fall below the 2.5% threshold for testing on an individual basis if both missing values were due to babies whose temperature was not taken within one hour.

3. **High performing units:** There were 58 NNU with no missing temperature records for 2013 or 2014 and where all babies had their temperature taken according to the NNAP standard – these are listed in Table 3. This list includes NNU with fewer than ten eligible babies.

Table 3: NNU where all eligible babies had temperature records, and all temperature-taking achieved the NNAP standard.

Unit name	Number of eligible babies
Airedale General Hospital	15
Basingstoke & North Hampshire Hospital	5
Bassetlaw District General Hospital	5
Bedford Hospital	13
Calderdale Royal Hospital	24
Chesterfield & North Derbyshire Royal Hospital	11
Colchester General Hospital	25
Countess Of Chester Hospital	10
Croydon University Hospital	32
Dewsbury & District Hospital	9
Dorset County Hospital	10
East Surrey Hospital	15
Epsom General Hospital	5
Frimley Park Hospital	19
Glangwili General Hospital	4
Guy's & St Thomas' Hospital	100
King's Mill Hospital	24
Leighton Hospital	25
Macclesfield District General Hospital	2
Pilgrim Hospital	7
Pinderfields General Hospital	31
Poole Hospital NHS Foundation Trust	18
Prince Charles Hospital	8
Princess Alexandra Hospital	18
Princess Anne Hospital	89
Princess Of Wales Hospital	4
Queen Alexandra Hospital	130
Queen Elizabeth Hospital, Gateshead	5
Queen Elizabeth Hospital, King's Lynn	12
Queen's Hospital, Burton On Trent	11
Rotherham District General Hospital	13
Royal Berkshire Hospital	18
Royal Derby Hospital	44
Royal Hampshire County Hospital	11
Royal Oldham Hospital	102

Royal Shrewsbury Hospital	29
Royal Surrey County Hospital	10
Royal United Hospital	18
Salisbury District Hospital	8
Scarborough General Hospital	4
Scunthorpe General Hospital	10
South Tyneside District Hospital	2
St Helier Hospital	15
St Mary's Hospital, Isle of Wight	8
Stoke Mandeville Hospital	24
Sunderland Royal Hospital	45
Tameside General Hospital	12
Tunbridge Wells Hospital	18
University Hospital Coventry	79
University Hospital Lewisham	49
Wansbeck General Hospital	5
Warrington Hospital	6
Watford General Hospital	21
West Cumberland Hospital	6
West Middlesex University Hospital	19
West Suffolk Hospital	9
Wexham Park Hospital	27
Worthing Hospital	6

Table 4: NNU with fewer than ten eligible babies: these were not assessed for outlier status

Unit name	**Number of eligible babies**
Basingstoke & North Hampshire Hospital	5
Bassetlaw District General Hospital	5
County Hospital, Staffordshire	6
Cumberland Infirmary	7
Dewsbury & District Hospital	9
Epsom General Hospital	5
Furness General Hospital	2
George Eliot Hospital	8
Glangwili General Hospital	4
Good Hope Hospital	7
Hinchingbrooke Hospital	5
Macclesfield District General Hospital	2
Nevill Hall Hospital	6
Pilgrim Hospital	7
Prince Charles Hospital	8

Princess Of Wales Hospital	4
Princess Royal Hospital	6
Queen Elizabeth Hospital, Gateshead	5
Queen Elizabeth The Queen Mother Hospital	6
Salisbury District Hospital	8
Scarborough General Hospital	4
South Tyneside District Hospital	2
St Mary's Hospital, Isle of Wight	8
University Hospital Of North Durham	6
Wansbeck General Hospital	5
Warrington Hospital	6
Warwick Hospital	7
West Cumberland Hospital	6
West Suffolk Hospital	9
Worthing Hospital	6
Yeovil District Hospital	4

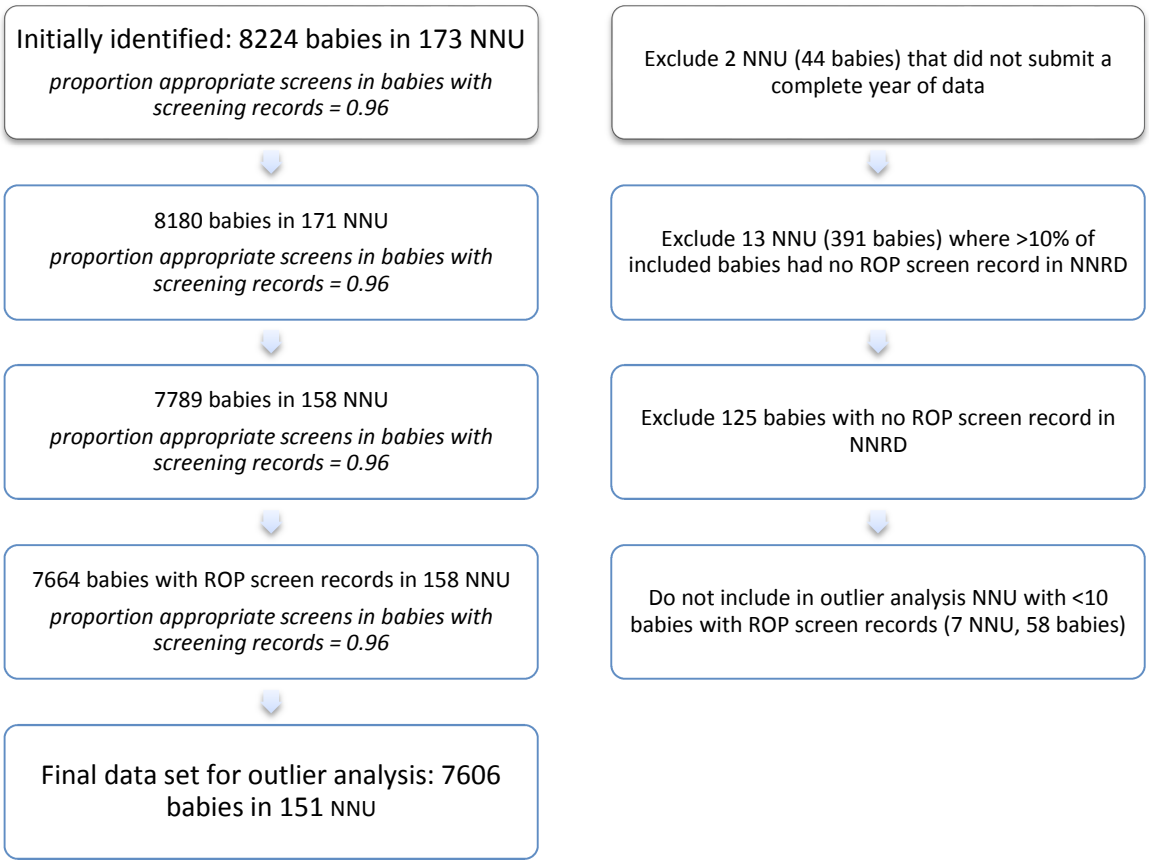
References:

1. Ohlssen DI, Sharples LD, Spiegelhlater DJ. (2007) A hierarchical modelling framework for identifying unusual performance in health care providers. Journal of the Royal Statistical Society A, 170: 865-890.
2. Walker K et al (2013) Public reporting of surgeon outcomes: low numbers of procedures lead to false complacency. The Lancet, 382: 1674-1677.

2014 Outlier analysis for the NNAP audit measure on: Babies <32 weeks or <1501g having Retinopathy of Prematurity (ROP) screening in accordance with current national guidelines.

1. Data set for outlier analysis

Figure 1. NNU and babies included in the ROP outlier analysis.



This report is based on data submitted to the National Neonatal Research Database (NNRD) by the 18th June 2015. An initial, provisional, analysis was done using data submitted before 9th March 2015, after which NNU were permitted to change their data.

For the NNU with small numbers of babies included in this question, we have very little power to detect large deviations from what is usual in the population. In other words, the probability of detecting a large true deviation is very low when the number of babies is small. Applying the outlier analysis to NNU with small numbers of babies may lead to false complacency and/or inability to detect good performance. Therefore we tested only the NNU with at least ten babies identified for inclusion in this audit question (see bottom right hand box in Figure 1). The NNU contributing fewer than ten babies were:

- Alexandra Hospital
- Bassetlaw District General Hospital
- Epsom General Hospital
- Furness General Hospital
- Harrogate District Hospital
- Princess Of Wales Hospital
- St Mary's Hospital, Isle of Wight

NNU where more than ten percent of included babies had no recorded screen are deemed non-participating due to lack of data. The 13 NNU whose ROP screening data is considered insufficient to participate in the outlier analysis are listed in Table 1. There has been a marked improvement in the recording of ROP screens since 2013, when 39 NNU were classified as non-participating.

Table 1 also shows the proportion of non-participating NNU that would achieve the NNAP standard if (1) all missing records were where the NNAP standard was not achieved (worst case) and (2) all missing records were where the NNAP standard was achieved (best case). We shall return to these after the outlier analysis in participating NNU.

Table 1 also shows that the majority of NNU with a large proportion of missing ROP screening records in 2014 had the same problem in 2013.

Table 1: Non-participating NNU, i.e. those with > 10% of eligible babies having no recorded ROP screen (table includes small NNU with missing ROP screening records)

Name	Number eligible babies	Number with no recorded screen	Proportion with recorded screen	Number with screen on time	Proportion of recorded screens on time	Worst case	Best case
Conquest Hospital	27	3	0.89	18	0.75	0.67	0.78
County Hospital, Staffordshire	13	3	0.77	10	1	0.77	1
Ealing Hospital*	21	6	0.71	15	1	0.71	1
Glan Clwyd Hospital*	31	4	0.87	26	0.96	0.84	0.97
Glangwili General Hospital*	22	4	0.82	18	1	0.82	1
Great Western Hospital*	48	27	0.44	13	0.62	0.27	0.83
James Cook University Hospital	82	10	0.88	65	0.9	0.79	0.91
King's Mill Hospital	34	5	0.85	19	0.66	0.56	0.71
Queen's Hospital, Romford**	44	5	0.89	28	0.72	0.64	0.75
South Tyneside District Hospital*	5	1	0.8	4	1	0.8	1
The Royal Free Hospital	17	5	0.71	10	0.83	0.59	0.88
Wrexham Maelor Hospital*	30	12	0.6	9	0.5	0.3	0.7
Ysbyty Gwynedd*	17	3	0.82	10	0.71	0.59	0.76

* Non-participating in 2013 report

** Low outlier in 2013 report

Response rates by gestational age category for non-participating and participating NNU are given in Table 2. In comparison with the same table from the 2013 report, there have been improvements in the rate of recording ROP screens in the non-participating units. As previously, the poorest recording is of screens for babies born after 32 weeks' gestation and weighing <1500g.

The non-participating NNU have a substantially lower rate of appropriate screening in babies with a recorded screen, with 255/303 (84%) babies with a recorded screen being screened on time, in contrast to 96% on time in the participating NNU. This is largely due to a high proportion (43/261=16%) of babies < 32 weeks being screened late (either pre or post discharge). Both recording and appropriate screening should be carefully scrutinized in the non-participating NNU, even though their data is considered insufficient to be included in the outlier analysis.

Table 2: ROP screening in non-participating and participating NNU by gestational age criterion

	ROP screening groups	Number babies	Babies with no screen recorded (% of eligible babies)	Babies with screen recorded	NNAP standard screen (% of screens recorded)	Screen early pre discharge	Screen early post discharge	Screen late pre discharge	Screen late post discharge	Screen recorded not achieving NNAP standard
Participant (158 NNU)	<32 weeks	6607	54 (0.8%)	6553	6350 (97%)	20	1	116	66	203 (3.1%)
	>= 32 weeks & <1501g	1182	71 (6.0%)	1111	1019 (92%)	27	9	7	49	92 (8.3%)
Non-participant (13 NNU)	<32 weeks	310	49 (16%)	261	217 (83%)	1	0	33	10	44 (17%)
	>= 32 weeks & <1501g	81	39 (48%)	42	38 (90%)	2	1	0	1	4 (10%)

2. Outlier analysis

Methods:

The data set used for the outlier analysis is shown in Figure 1. The proportion appropriately screened in each participating NNU was based on babies with screening records only – (where none of the participating NNU had more than ten percent of their screening records missing).

The first step in our staged methodology to identify potentially unusual NNU is an initial screen. In this step we use all the data (described in the above paragraph) to fit a model for appropriate ROP screening in the population. The model we use is a random effects model. As indicated in Figure 2, we then use the model's 5% prediction limit to flag NNU with low rates of appropriate screening. We would expect to see 5% of NNU below the 5% threshold due to chance.

However, the first stage model is based on all the NNU, including those that may be unusual. The second stage of analysis comprises further testing of the NNU, by fitting a second model for ROP screening in the NNU that were not identified as unusual at the first stage. Using the second stage model, which allows smaller variation among NNU to be considered usual, we identify in Figure 3 potentially unusual NNU using stricter limits: (alert)

2.5% and (alarm) 0.1%. The resulting funnel plot is useful for individual NNU to check their performance.

The final stage is to adjust for multiple testing. The final funnel plot is used by NNAP to identify possibly unusual NNU, allowing for the fact that 151 tests are made, and with each test the probability of incorrectly identifying a NNU as unusual is 0.025, so that without adjustment the probability of incorrectly identifying one of the NNU in the analysis as unusual is high at approximately 98%. Here we set the false discovery rate (that is the expected proportion of false positives) to 10%.

We note that the thresholds we have used are decided *a priori*, and are consistent with those used by national audit programmes, including past NNAP analyses, and in the literature on identification of unusual performance. Especially for NNU close to a threshold, the choice of the cut-off may determine if the NNU is flagged or not. We emphasise that this process is useful to flag potentially unusual performance, but further non-statistical investigations of flagged NNU are necessary.

Results:

NNU below threshold: We identified nine NNU below the 5% threshold at the initial stage.

Figure 2 is a funnel plot indicating these NNU with the 5% threshold for the initial stage.

Table 3 also indicates with asterisks the NNU that were below the 2.5% threshold at stage 2, before adjustment for multiple testing. One NNU not flagged at the initial stage was flagged at the second stage (with 103 babies, 93 (90%) screened on time).

The accompanying funnel plot (Figure 3) is useful for individual NNU to compare their performance to the overall rate in the population. For NNAP purposes, adjustment for multiple testing sets a wider interval to identify unusual performance.

Figure 2: Funnel plot for proportion of babies with screening records who were screened according to NNAP standards. The 5% control limit is shown, and NNU identified as possibly unusually low are highlighted. These NNU are not included in the final model for usual performance in the population.

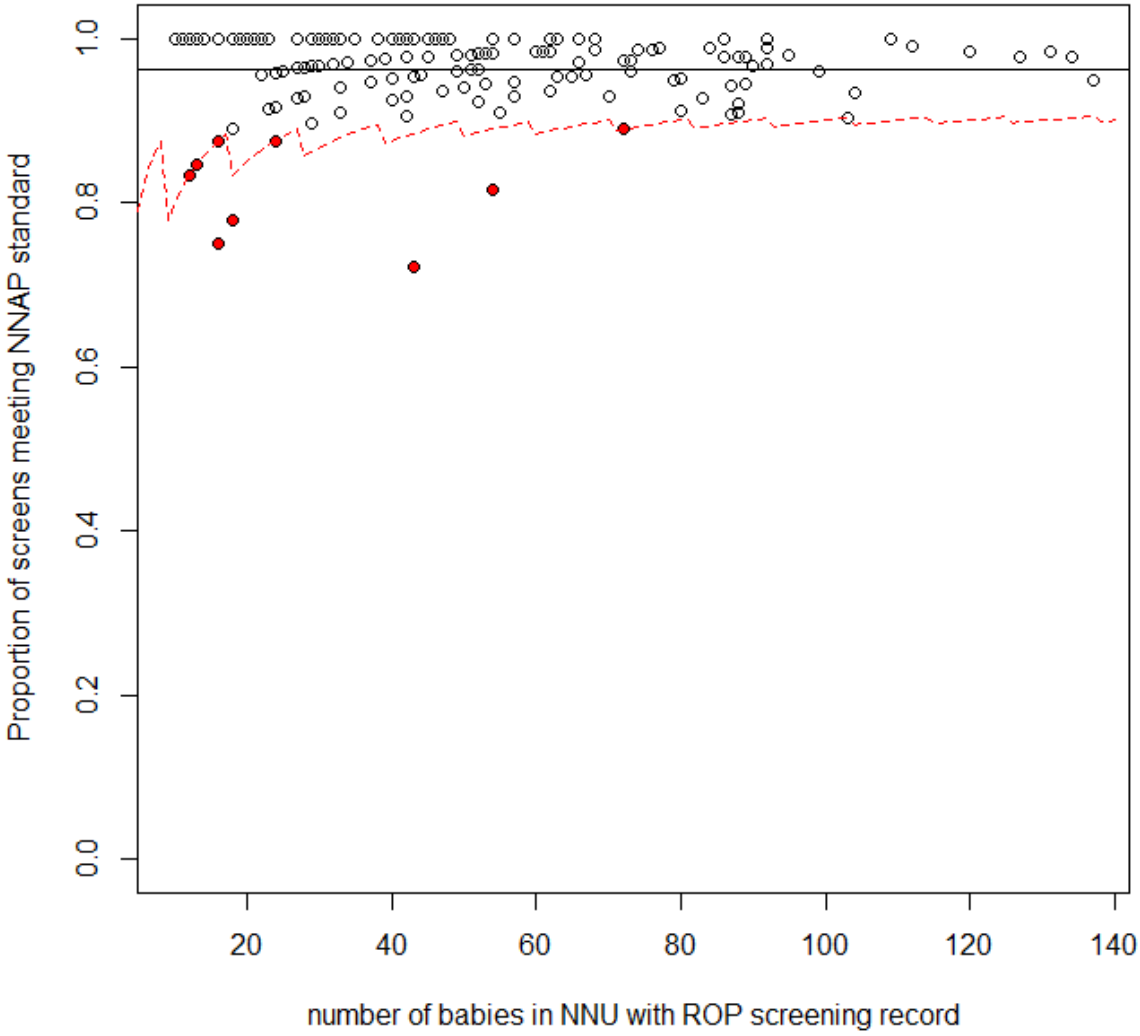
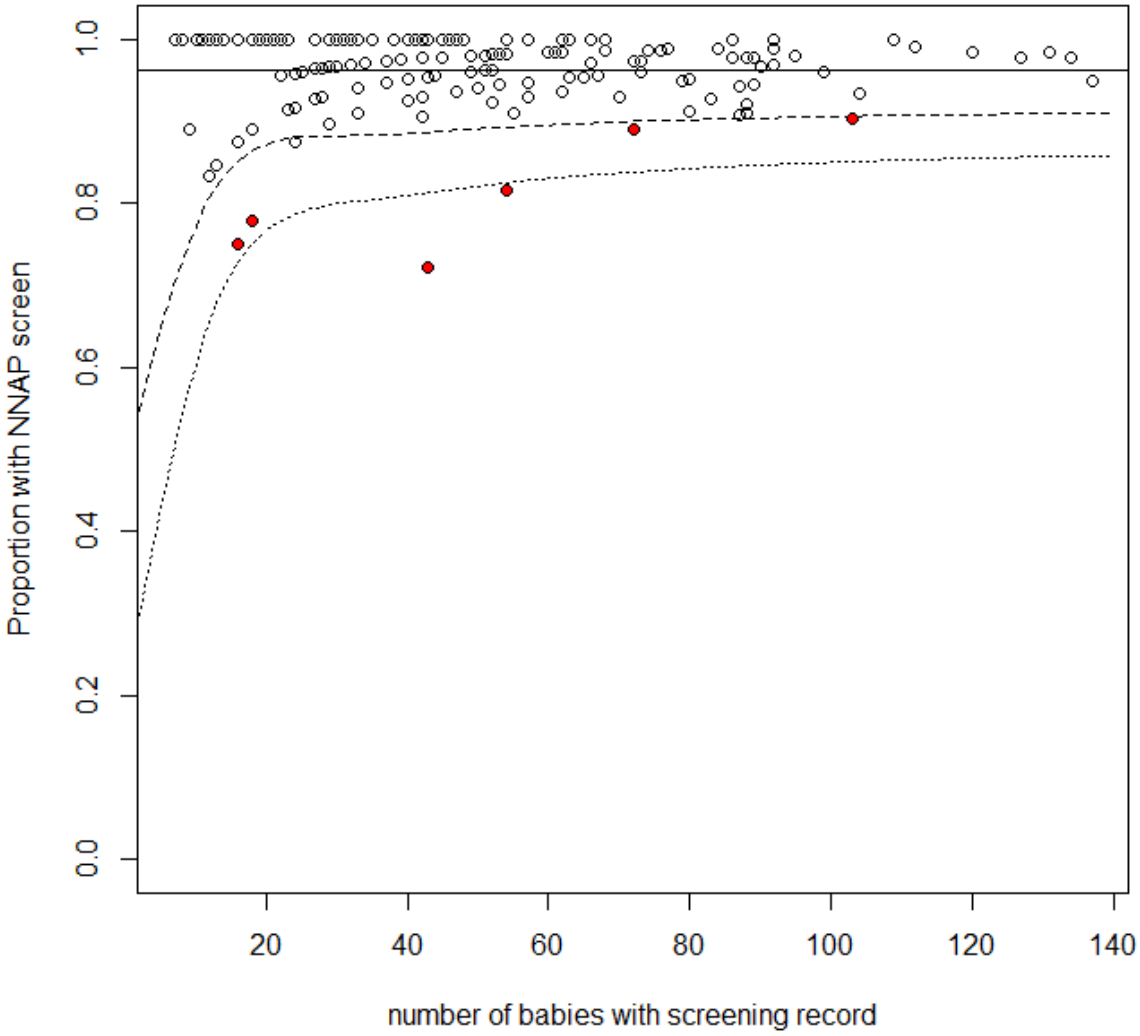


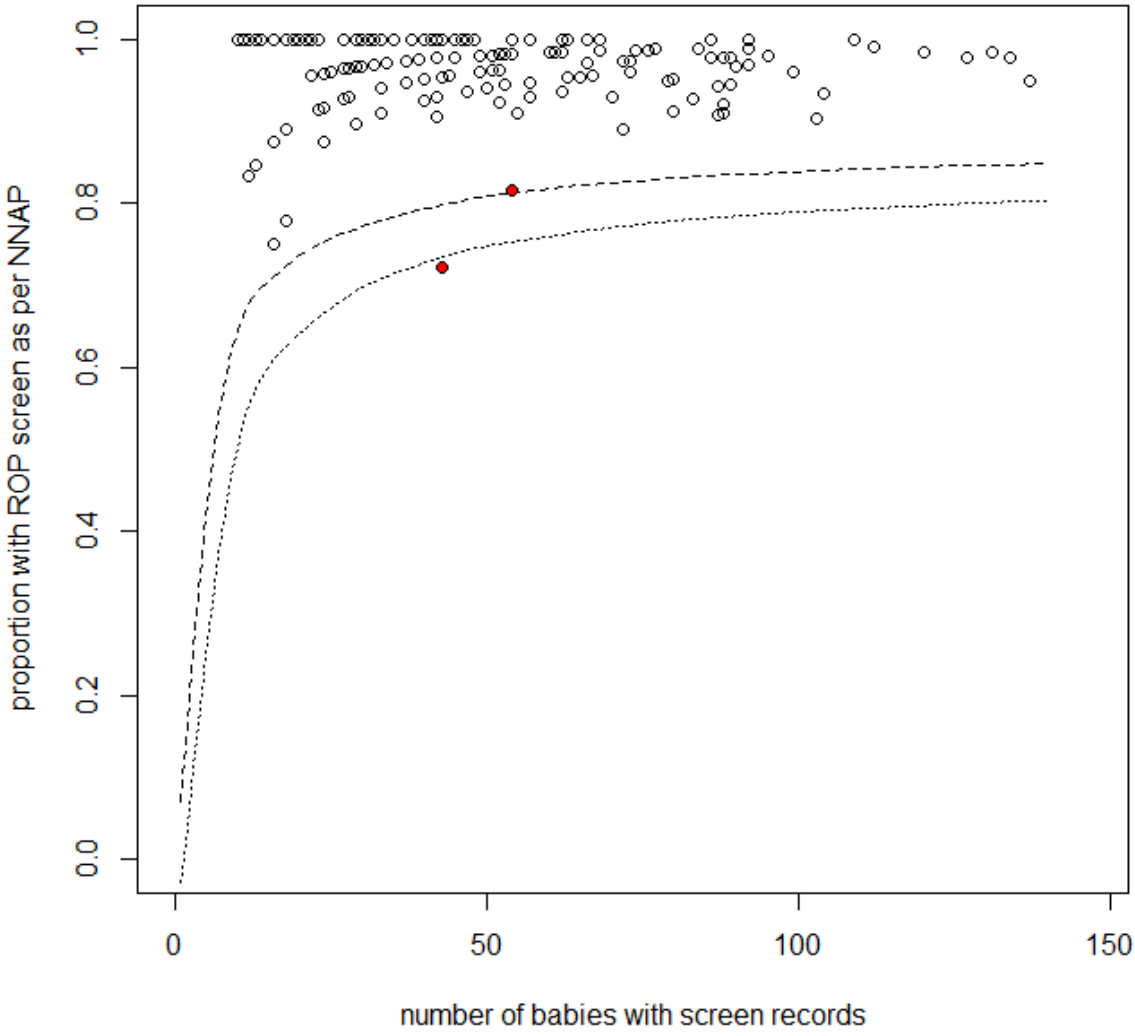
Figure 3: Funnel plot for proportion of babies with screening records who were screened according to NNAP standards. The smoothed 2.5% and 0.1% control limits for the stage 2 model are shown, and NNU identified as possibly unusually low on an individual basis are highlighted.



Confirmation of 2014 low-outliers

The results of setting the false discovery rate to 10% are shown in the following Figure 4. **Two NNU are indicated as unusual (low outliers)** – **University Hospital of North Tees** with 31/43 (72%) screens on time and **Queen Elizabeth Hospital, Woolwich** with 44/54 (81%) screens on time. For both these NNU, this is explained by late screens (see Table 3).

Figure 4: Funnel plot for proportion of babies with screening records who were screened according to NNAP standards. The smoothed 2.5% and 0.1% control limits shown, with false discovery rate set to 10%, and NNU identified as possibly unusually low are highlighted.



Sensitivity analysis for participating NNU with missing ROP screens:

Participating NNU all have less than ten percent missing records (most have very little missing) and therefore are unlikely to fall below the alert threshold even if all their missing screens all in fact fail to meet the NNAP standard. We checked all participating NNU, and found that no NNU would fall below the alert threshold in this worst case scenario. We also checked the two NNU that do fall below the thresholds and found that even in the best possible case, where all their missing values were good ROP screens, they would still fall below these thresholds. This is not surprising as **University Hospital of North Tees** had only one missing value and **Queen Elizabeth Hospital, Woolwich** had none missing.

Non-participating NNU:

We now return to the non-participating NNU listed in Table 1. It is clear that in the worst case all, and in the best case many of these would fall below the "alert" threshold for individual testing indicated in Figure 3. Adjusting for multiple testing, the following non-participant NNU would be identified as unusual in the worst case: **The Royal Free Hospital**, **Ealing Hospital, Ysbyty Gwynedd, Queen's Hospital, Romford, Great Western Hospital, James Cook University Hospital, Conquest Hospital, Wrexham Maelor Hospital**, and in the best case: **Queen's Hospital, Romford** and **Wrexham Maelor Hospital**.

3. **High performing units**: There were 40 NNU with no missing screening records and where all babies were screened according to the NNAP standard – these are listed in Table 4. This is an improvement on the outcome in 2013, when 22 NNU achieved this standard.

Table 4: NNU with all eligible babies screened, and all screens achieving the NNAP standard.

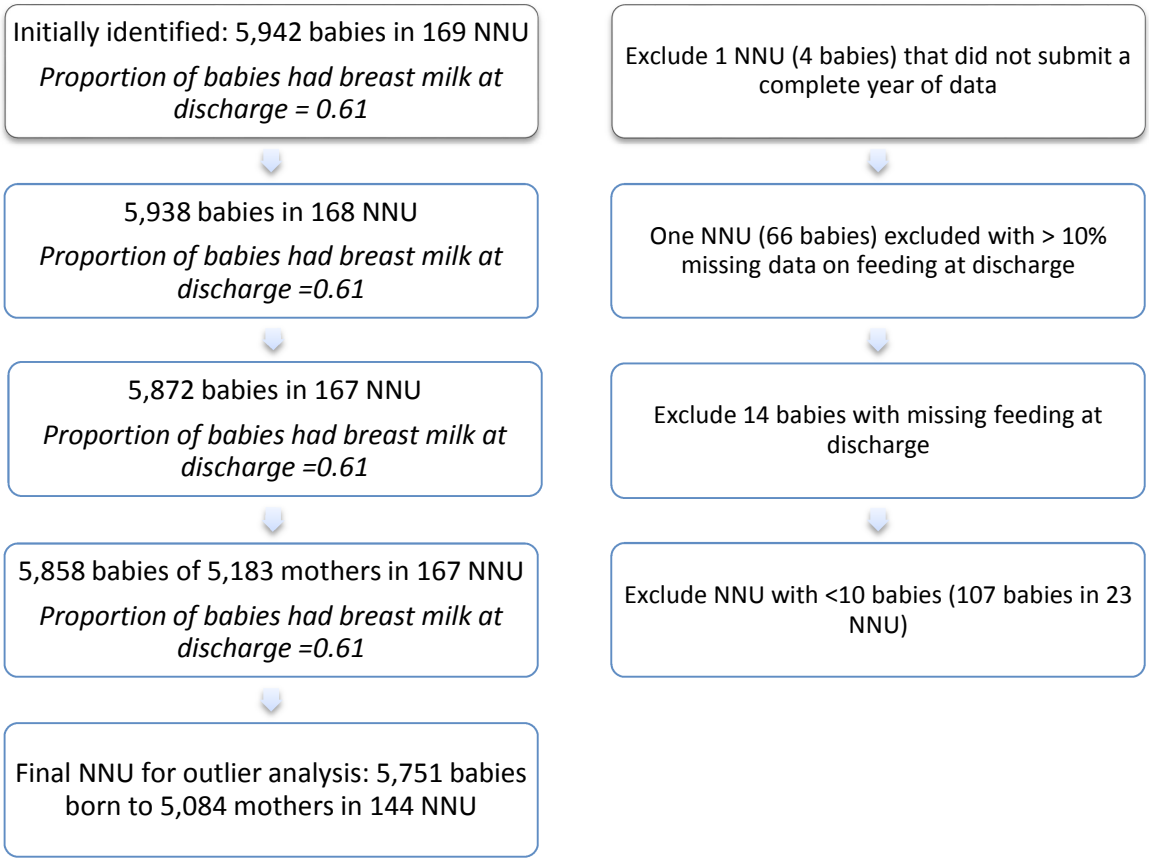
Unit Name	Eligible babies
Arrowe Park Hospital	54
Barnsley District General Hospital	30
Basingstoke & North Hampshire Hospital	21
Bassetlaw District General Hospital	8
Bedford Hospital	16
Bradford Royal Infirmary	109
Calderdale Royal Hospital	46
Croydon University Hospital	68
Doncaster Royal Infirmary	38
East Surrey Hospital	41
Epsom General Hospital	8
George Eliot Hospital	13
Harrogate District Hospital	7
Ipswich Hospital	27
Leighton Hospital	29
Macclesfield District General Hospital	13
Nevill Hall Hospital	23
North Devon District Hospital	13
Oxford University Hospitals, Horton Hospital	12
Princess Alexandra Hospital	33
Queen Charlotte's Hospital	66
Queen's Hospital, Burton On Trent	31
Royal Glamorgan Hospital	21
Royal Shrewsbury Hospital	48
Russells Hall Hospital	47
Salisbury District Hospital	19
Scarborough General Hospital	11

St Mary's Hospital, Isle of Wight	8
St Richard's Hospital	20
Sunderland Royal Hospital	45
Tameside General Hospital	23
Torbay Hospital	14
University Hospital Of North Durham	22
University Hospital Of North Staffordshire	92
University Hospital Of South Manchester	57
Warrington Hospital	19
Warwick Hospital	18
West Cumberland Hospital	10
Yeovil District Hospital	11
York District Hospital	32

2014 Outlier analysis for the NNAP audit measure on: Babies <33 weeks gestation receiving any of their mother's milk when discharged from a neonatal unit to home.

1. *Data set for outlier analysis*

Figure 1. NNU and babies included in the outlier analysis.



This report is based on data submitted to the National Neonatal Research Database (NNRD) before the 18th June 2015. An initial analysis was done using data submitted before 9th March 2015, after which NNU were permitted to change their data. Discharge feeding was identified from feeding records on the day of discharge, and if this was missing, the feeding record on the previous day.

Figure 1 illustrates the process by which the NNU to be included in the analysis were selected.

One NNU (**Withybush Hospital**) was excluded as it did not submit a full year of data.

Non-participating NNU:

One NNU, **Liverpool Women's Hospital**, was deemed to be non-participating as it had 61/66 (92%) missing discharge feeding records.

There were 23 NNU with fewer than ten babies eligible for inclusion in this question, and these are listed in Table 4 at the end of this document. We note that 51/107 (48%) of babies in these small NNU were fed with some mother's milk at discharge, which is lower than the

proportion in the population as a whole (61%) – however with such small numbers of babies in each NNU, we do not have power to detect unusual performance at individual NNU level.

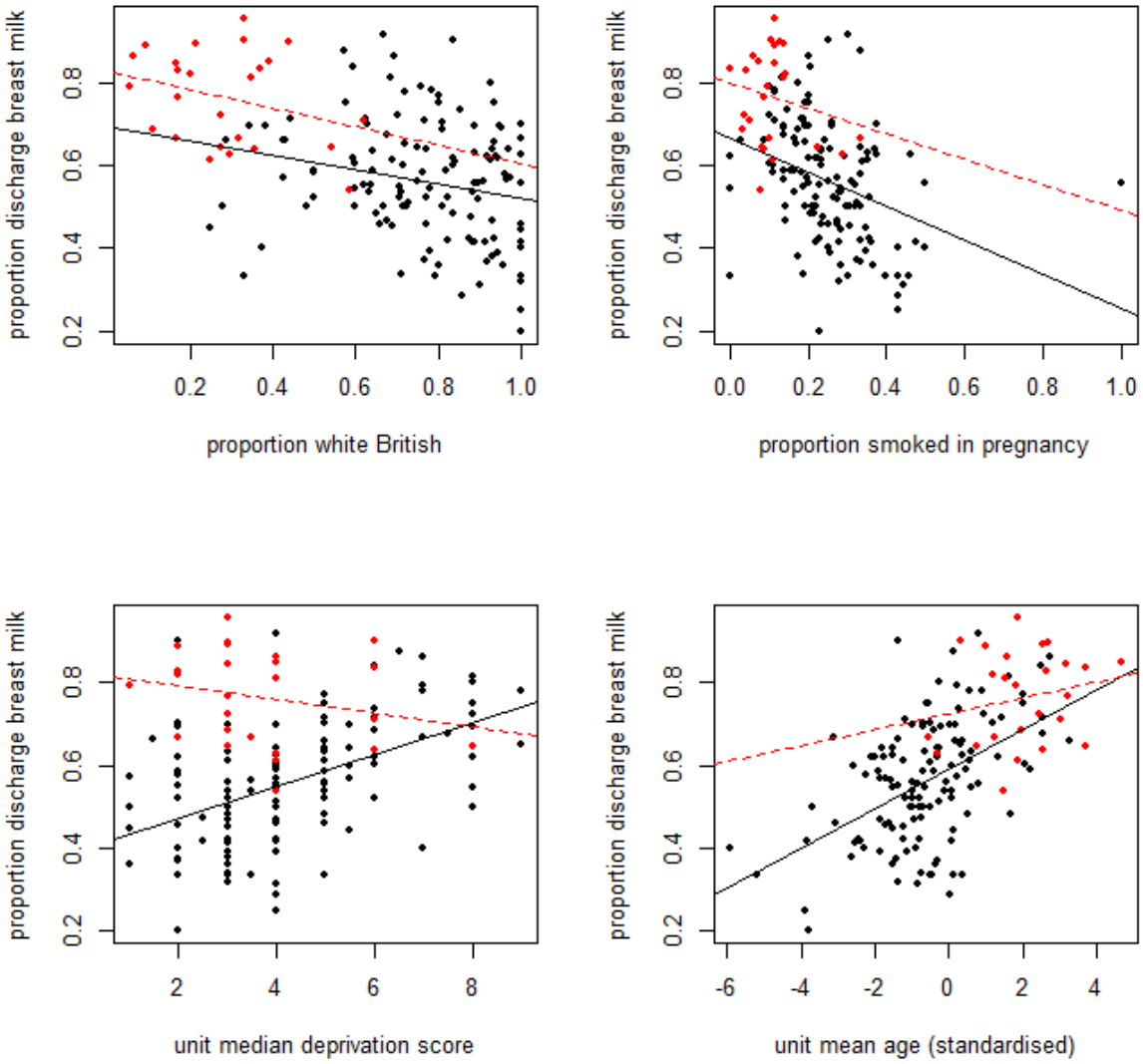
2. Risk adjustment

Characteristics of the mother are known to affect the probability she will feed her preterm babies breast milk – these characteristics also vary between areas and therefore between NNU. We wished to compare the actual number of babies breast milk fed at discharge with the number we would expect to be breast milk fed in each NNU, given the characteristics of their mothers. We used a statistical model (hierarchical logistic regression model) to estimate the expected probability each mother in the population breast milk feeds at discharge, based on her characteristics.

Variables we included in the model were mother's age, the square of mother's age, whether the birth was a multiple, mother's ethnicity ("white-British" versus all other categories, as per the standard NHS ethnicity categories)[1], smoking during pregnancy (any versus none), and decile of area-level ranked multiple deprivation (IMD) score based on mother's postcode.[2] As the multiple deprivation scores for Wales and England are not directly comparable, we adjusted the Welsh deprivation scores to the English deprivation score scale using a regression method.[3] To allow for known differences in the association between socioeconomic factors and breast-feeding in London compared to the rest of England and Wales, we included an indicator for London and an interaction between area-level deprivation and residence in London.[4] Allowance for differences between London and the rest of the country was not made in the model presented for the initial report based on data submitted up to March. As some of the variables that we used in the model were partly missing (i.e. not recorded in the NNRD), we simultaneously fitted models to impute maternal age, smoking, ethnicity and decile of deprivation score if these were missing. We did this by assuming prior distributions for these variables based on the observed distribution in the population.

Figure 2 illustrates the broad pattern of the observed associations between sociodemographic factors and breast milk feeding at discharge at NNU level. London NNU, highlighted in red, are clearly different to the rest of the country, but the associations between the chosen factors and breast milk feeding are quite similar (red and black lines have similar slopes) with the exception of deprivation score, where breast milk feeding on average increases with increasing area-level advantage across the rest of the country, but there is a small negative association in London.

Figure 2: Scatterplots showing NNU-level proportion breast milk feeding at discharge versus (a) proportion white British, (b) proportion who smoked in pregnancy, (c) median deprivation score and (d) mean age; with fitted linear regressions. London is represented in red and the rest of England and Wales in black.



Eleven percent of mothers were missing smoking status, 12% missing ethnicity, 4% missing deprivation scores and <1% missing age. The parameter estimates for the logistic regression model, which give the predicted change in the log odds of breast milk feeding for a specified change in the predictor, are given in Table 1. For factor variables, (e.g. smoking during pregnancy, British ethnicity) the interpretation is relative to the baseline, for instance for women who smoked during pregnancy, compared to women who did not smoke during pregnancy, the expected log odds of breast milk feeding at discharge on average decreased by 1.10. Converting to odds, this means that on average the odds of breast milk feeding in smoking mothers was about 1/3 that of breast milk feeding in non-smoking mothers. For continuous variables the interpretation is per unit change – for example, per year of age. The parameter estimates show that the odds of breast milk feeding is negative quadratic in age (the pattern is an average increase with age that plateaus off for the oldest mothers).

It is lower for smokers than non-smokers and “white-British” mothers versus any other ethnic category, and lower for mothers of multiples than mothers of singletons, and is higher in London than the rest of England and Wales. Overall the log-odds of breast milk feeding increases as area-level socioeconomic deprivation index increases, i.e. in areas of greater socioeconomic advantage, but there is little evidence of association between socioeconomic advantage and breast milk feeding in London. As we estimated this model in a Bayesian framework, we give 95% credible intervals for the log odds – the probability that the log odds ratio lies between these bounds is 0.95.

Table 1: Parameter estimates from logistic model for breast milk feeding at discharge.

	Log odds	95% credible interval	
Intercept	0.70	0.52	0.88
Age	0.050	0.040	0.060
Age squared	-0.0058	-0.0070	-0.0046
Smoking during pregnancy	-1.10	-1.26	-0.94
Deprivation (decile)	0.14	0.12	0.17
British ethnicity	-0.67	-0.82	-0.52
Multiple birth	-0.24	-0.38	-0.11
London	1.21	0.83	1.59
London X Deprivation	-0.15	-0.23	-0.07
Random intercept			
	Estimate	Posterior SD	
SD	0.3258	0.048	

We used the model described above to calculate the probability that each mother fed with breast milk at discharge, and summed these probabilities for babies within mothers within NNU to calculate the expected number of babies breast milk fed at discharge in each NNU. We then compared the actual (or observed) number of babies breast milk fed with the expected number by calculating the odds ratio of observed to expected breast feeding for each NNU. If this value is close to one, the number of babies breast milk fed is as we expect, given the characteristics of the mothers we included in the adjustment model. If this value is much smaller (or much greater) than one, then the NNU has a smaller (greater) number of babies breast milk fed than we expect. What is considered “small” and “great” enough to be unusual depends on the distribution of the ratio in the population of *usual*, i.e. not *unusual*, NNU.

3. Outlier analysis

Methods:

To answer the question "what is the evidence that any NNU perform worse than we would expect the true worst in the population of *usual* NNU to be" we used a staged methodology to robustly identify potentially unusual NNU, adapted from Ohlssen et al.[5] The first stage is to estimate a model for the log odds ratio of observed to expected breast milk feeding based on data from all NNU. While the risk adjustment explained some of the variability between NNU, we used a random effects model to allow for additional variability. For each NNU we tested the hypothesis that the observed odds of breast milk feeding was consistent with the expected odds of breast milk feeding, versus the alternative that it was smaller than expected. To test these hypotheses we derived one-sided p-values for each NNU, based on the model standard errors appropriate to identify outliers.[6] The p-values give the probability of observing our data (or more extreme data) for each NNU if the observed odds of breast feeding is as expected. We chose the thresholds 5% and 95% to identify NNU at the first stage of testing. We illustrated this using a funnel plot (figure 3).[7]

However, by basing the model on all NNU we may include unusual NNU in the model of what is usual in the population. The second stage is re-estimation of the model without the influence of the potentially unusual NNU flagged at the first stage; and hence to derive robust estimates of the p-values for all NNU. At the second stage we used the standard "alert" limits of 2.5% and 97.5% (and also show the "alarm" 0.1% and 99.9% limits in funnel plots). This analysis is useful for individual NNU to compare their performance to that of other NNU (Figure 4).

At the final stage we adjusted for multiple testing by setting the expected number of false positive tests is set to 10%.[8] This is the basis for the NNAP identification of outliers, with adjustment for multiple testing (figure 5).

Results:

At the first stage, based on control limits of 5% and 95%, we identified three NNU with potentially lower than expected odds of breast milk feeding at discharge. We also identified thirteen NNU with potentially higher than expected odds of breast milk feeding at discharge. The results of the first stage are also illustrated by a funnel plot in Figure 3.

There were three NNU below the 2.5% threshold at the second stage of testing, and one of these was below the 0.1% threshold. There were six NNU above the 97.5% threshold at stage 2, but none above the 99.9% threshold. The results of the second stage of testing also illustrated by a funnel plot in Figure 4.

Figure 3: First stage funnel plot of the odds ratio of actual breast milk feeding to expected breast milk feeding. The horizontal axis is the precision. Without risk adjustment (as for the other NNAP outlier analyses) the precision is the size of the NNU; however when we use a logistic model to predict the outcome, and compare observed rates with what we predict from the model, the precision needs to take this into account. The precision for an NNU is now equal to the inverse of the estimated variance of its log odds ratio of observed to expected babies breast fed. Thresholds for the first stage are 5% and 95% prediction limits.

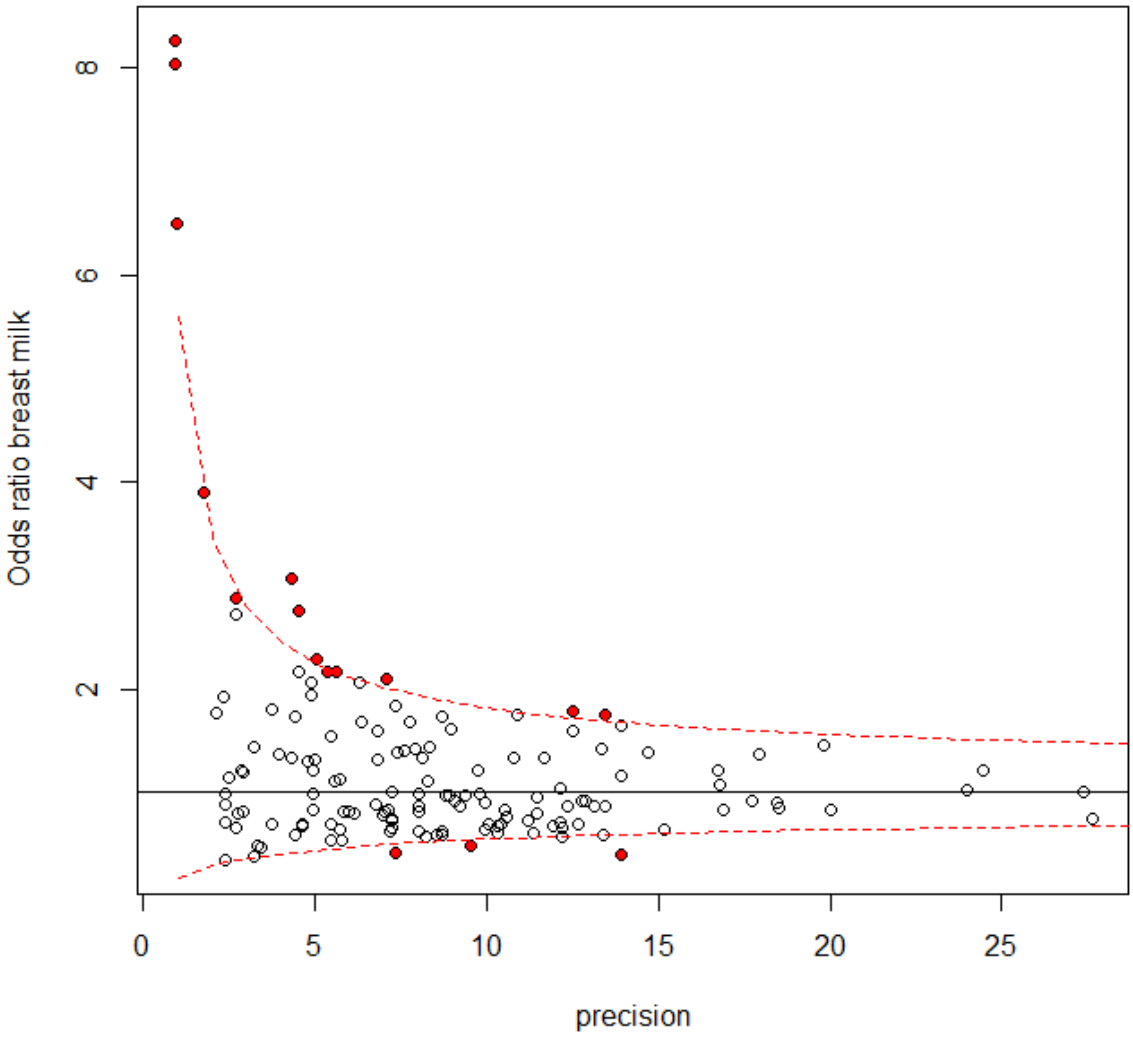


Figure 4: Second stage funnel plot of the odds ratio of actual breast milk feeding to expected breast milk feeding. NNU identified beyond the second stage "alert" thresholds are highlighted. Thresholds are "alert" 2.5% and 97.5%, and "alarm" 0.1% and 99.9% prediction limits.

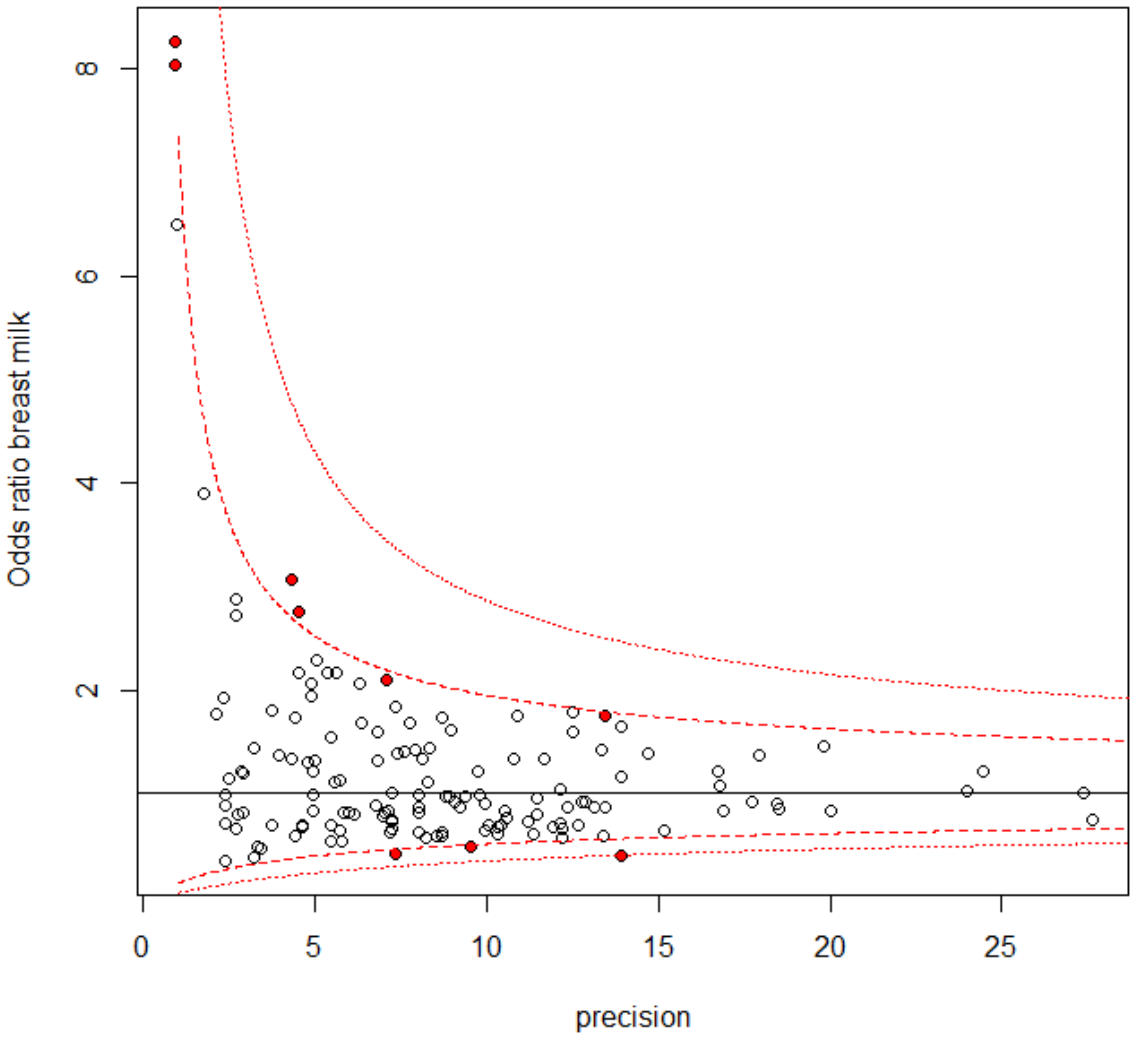
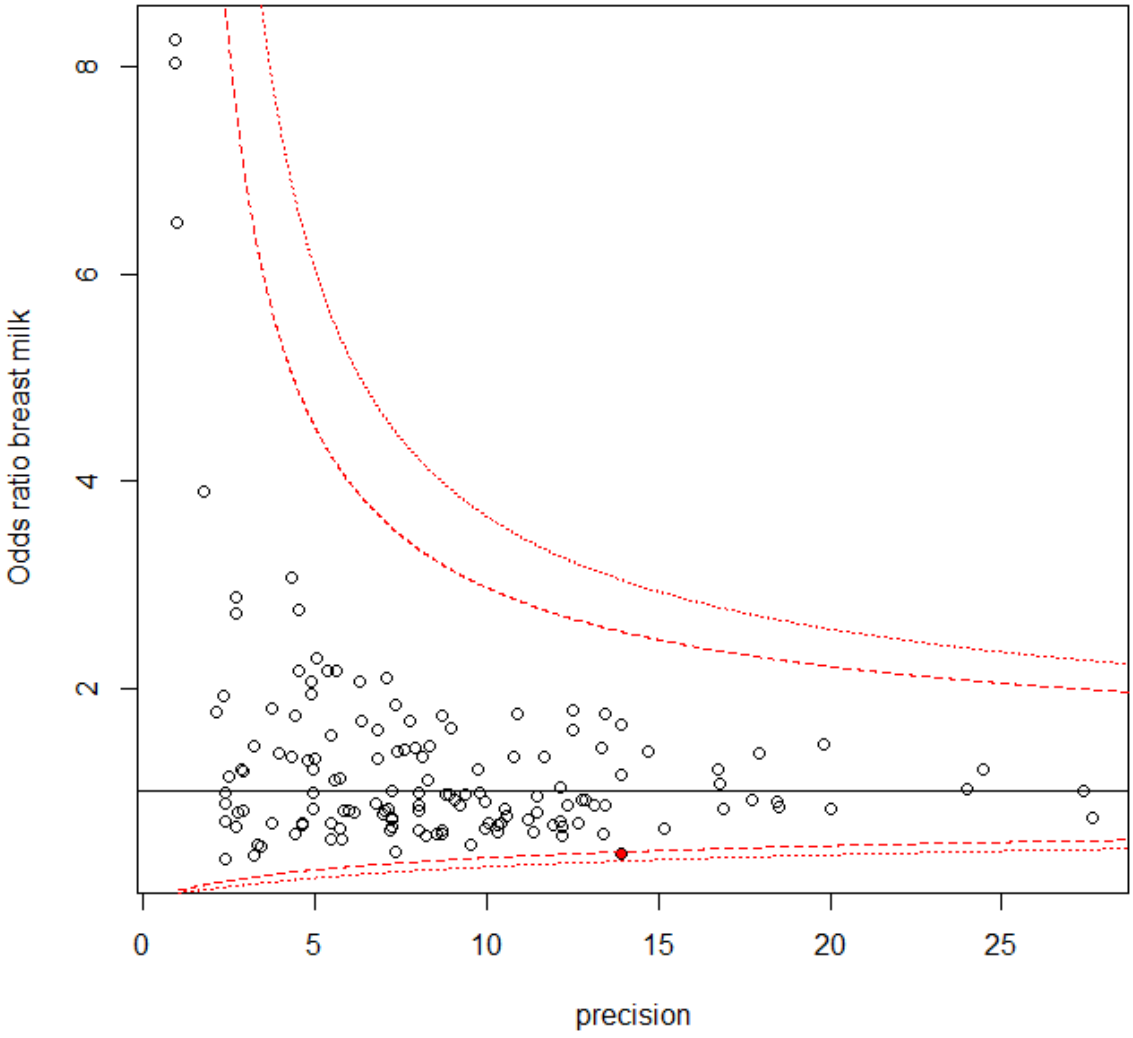


Figure 5: Final stage funnel plot of the odds ratio of actual breast milk feeding to expected breast milk feeding. One NNU identified after adjustment for multiple testing is highlighted. Thresholds are for 10% and 1% false discovery rates.

Confirmation of 2014 low-outliers

Medway Maritime Hospital is identified as a low outlier at the final stage, setting the false discovery rate to 10%. The final stage of testing is illustrated in Figure 5. No other NNU are identified as unusual at this stage.



We did not do a sensitivity analysis for missing responses as with the exception of the one NNU excluded from the analysis, there were too few missing to have an impact (14/5872).

Table 4 NNU with fewer than ten eligible babies not included in the outlier analysis.

Unit Name	Number of eligible babies
Alexandra Hospital	1
Bassetlaw District General Hospital	3
Bedford Hospital	6
County Hospital, Staffordshire	5
Ealing Hospital	6
Furness General Hospital	1
Glangwili General Hospital	8
Harrogate District Hospital	1
Hereford County Hospital	5
Oxford University Hospitals, Horton Hospital	1
Princess Of Wales Hospital	4
Princess Royal Hospital	2
Queen Elizabeth Hospital, Gateshead	9
Scarborough General Hospital	1
South Tyneside District Hospital	1
St Mary's Hospital, Isle of Wight	7
The Royal Free Hospital	2
Torbay Hospital	8
University Hospital Of North Durham	9
Warwick Hospital	8
West Cumberland Hospital	9
West Suffolk Hospital	9
Yeovil District Hospital	1

References

1 NHS ethnicity categories, at http://www.datadictionary.nhs.uk/data_dictionary/attributes/e/end/ethnic_category_code_de.asp?shownav=1

2 Index of multiple deprivation at https://www.gov.uk/government/collections/english-indices-of-deprivation

3 Payne RA, Abel GA (2012). UK indices of multiple deprivation: a way to make comparisons across constituent countries easier. *Health Stat Q* **53**: 22-37.

4 Oakley LL et al (2013). Factors associated with breastfeeding in England: an analysis by primary care trust. *BMJ Open* **3**: e002765.

5 Ohlssen DI Sharples LD Spiegelhalter DJ (2007). A hierarchical modelling framework for identifying unusual performance in health care providers. *Journal of the Royal Statistical Society A*, **170**: 865-890.

6 Jones HE Spiegelhalter DJ (2011). The identification of “unusual” health-care providers from a hierarchical model. *The American Statistician* **65**: 154-163.

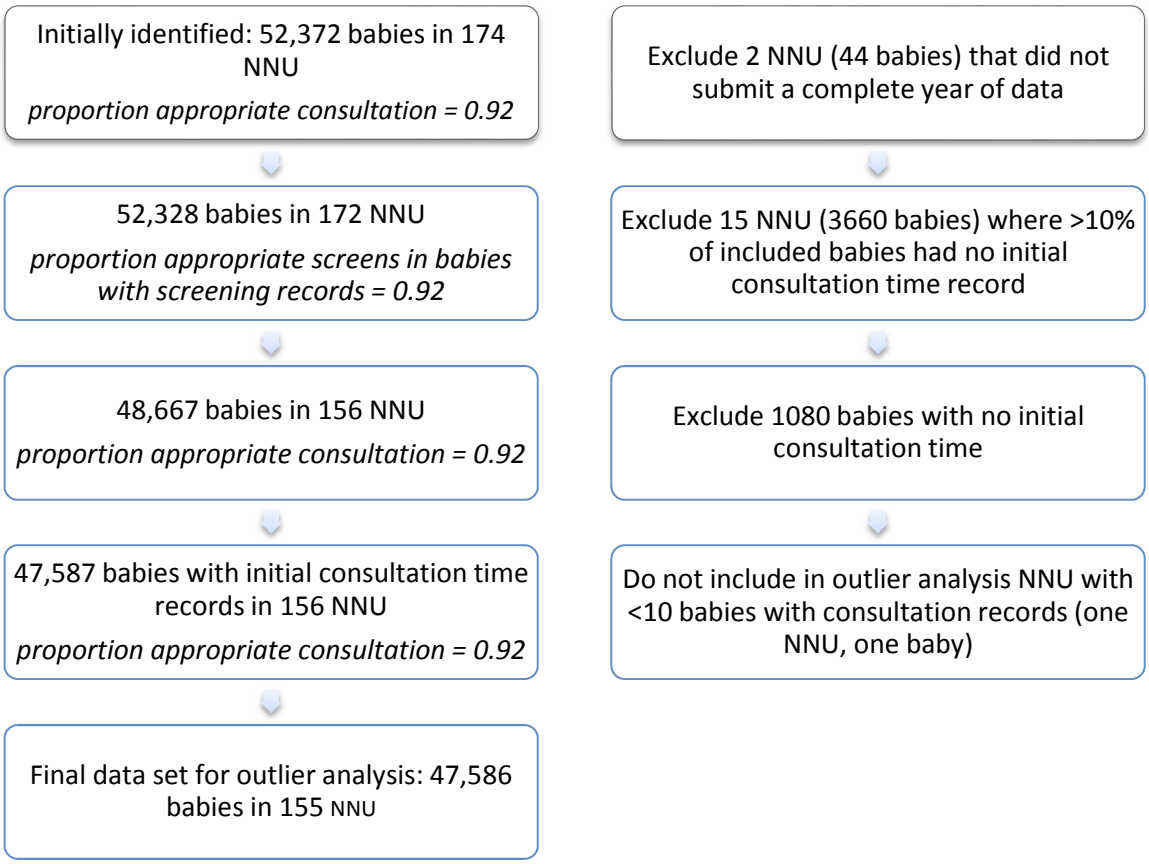
7 Spiegelhalter DJ (2005) Funnel plots for comparing institutional performance. *Statistics in Medicine* **24**:1185-1202.

8 Jones HE Ohlssen DI Spiegelhalter DJ (2008) Use of the false discovery rate when comparing multiple health care providers. *Journal of Clinical Epidemiology*, **61**: 232-240.

2014 Outlier analysis for the NNAP audit measure on: Consultation with parents by a senior member of the neonatal team within 24 hours of admission

1. Data set for outlier analysis

Figure 1. NNU and babies included in the outlier analysis.



This report is based on data submitted to the National Neonatal Research Database (NNRD) before the 18th June 2015. An initial analysis was done and reported to NNU using data submitted before 9th March 2015, after which NNU were permitted to change their data.

Two NNU, **Withybush Hospital** and **King George Hospital**, did not submit a complete year of data and therefore were not included in the outlier analysis.

There were 15 NNU whose consultation time records were considered insufficient to participate in the outlier analysis, as they had no first consultation records for >10% of eligible babies, and these are listed in Table 1. There has been a marked improvement in the recording of consultation time since 2013, when 39 NNU were classified as non-participating. There was a further improvement in the 2015 data amendment period - there were 22 NNU classified as non-participating in the initial report based on data submitted before 9th March 2015. Eleven of the 15 non-participating NNU for the 2014 outlier analyses were also non-participating in the 2013 data outlier analyses for this question.

There was one NNU, **Bronglais General Hospital**, which contributed only one baby and therefore was not included in the outlier analysis.

Table 1: Non-participating NNU, having > 10% of eligible babies with no record in the database of an initial consultation time

Unit Name	Number eligible babies	Number with no recorded initial consultation time	Proportion with recorded initial consultation time	Number on time	Proportion of recorded consultations on time	Worst case	Best case
Alexandra Hospital*	99	12	0.88	78	0.9	0.79	0.91
Glan Clwyd Hospital*	192	47	0.76	108	0.74	0.56	0.81
Glangwili General Hospital	123	27	0.78	77	0.8	0.62	0.85
Great Western Hospital*	276	107	0.61	136	0.8	0.49	0.88
Nottingham City Hospital*	316	35	0.89	247	0.88	0.78	0.89
Nottingham University Hospital (QMC)*	322	34	0.89	266	0.92	0.83	0.93
Princess Of Wales Hospital*	200	29	0.86	145	0.85	0.73	0.87
Princess Royal University Hospital	255	30	0.88	182	0.81	0.71	0.83
Queen Charlotte's Hospital*	385	48	0.88	287	0.85	0.75	0.87
Royal United Hospital*	466	55	0.88	325	0.79	0.69	0.82
South Tyneside District Hospital*	63	9	0.86	47	0.87	0.75	0.89
Stepping Hill Hospital*	212	36	0.83	125	0.71	0.59	0.76
The Royal Free Hospital	199	37	0.81	107	0.66	0.54	0.72
University Hospital Of Wales	397	59	0.85	274	0.81	0.69	0.84
Wrexham Maelor Hospital*	155	39	0.75	106	0.91	0.68	0.94

* Also non-participating in 2013.

2. Outlier analysis

Methods:

The data set used for the outlier analysis is shown in Figure 1. The proportion of babies whose parents had a consultation within 24 hours in each participating NNU was based on babies with consultation time records only (where none of the participating NNU have more than ten percent of their consultation times missing).

We used a staged methodology to identify unusual performance, adapted from Ohlssen et al.[1] The first step is an initial screen. In this step we used all the data from participating NNU to fit a model for appropriate consultation in the population (figure 2). As the variability in the rates of consultation at NNU level is larger than we would expect based on within-NNU sampling error, we used a random effects model, which allows for some variability between NNU. We then used the first stage model's 5% prediction limit to flag NNU with low rates of appropriate consultation.

The first stage model is based on all the NNU, including those that may be unusual. The second stage of analysis comprises further testing of the NNU, by fitting a second model for on-time consultation in the NNU that were not identified as unusual at the first stage. Using the second stage model, which allows smaller variation among NNU to be considered usual, we identify potentially unusual NNU using the standard limits used by NNAP: (alert) 2.5% and (alarm) 0.1%. The resulting funnel plot is useful for individual NNU to compare their performance to that of other units (figure 3).

The final stage is to adjust for multiple testing. The final funnel plot (figure 4) is used by NNAP to identify NNU as "outliers", allowing for the fact that 155 tests are made, and with each test the probability of incorrectly identifying a NNU as unusual is 2.5%, so that without adjustment the probability of incorrectly identifying a NNU as unusual is approximately 98%. To avoid identifying a large number of false positives, we control the expected proportion of false positives (this is known as the false discovery rate). We illustrate this using a funnel plot that shows the thresholds for false discovery rates 10% and 1%.

We note that the thresholds we have used are decided *a priori*, and are consistent with those used by national audit programmes, including past NNAP analyses, and in the literature on identification of unusual performance. Especially for NNU close to a threshold, the choice of the cut-off may determine if the NNU is flagged or not. We emphasise that this process is useful to flag potentially unusual performance, but further non-statistical investigations of flagged NNU are necessary.

Results:

NNU below thresholds at stages 1 and 2: We identified six NNU below the 5% threshold at the initial screening stage. The corresponding funnel plot is shown in Figure 2.

At the second stage, we identified the same six NNU below the 2.5% prediction limit of the stage 2 model, and one further NNU not identified at the first stage. One NNU was below the 0.1% threshold for the stage 2 model.

The results of the stage 2 analysis are illustrated in Figure 3.

Figure 2: Funnel plot for proportion of babies with initial consultation time records who had a consultation within 24 hours of birth. The 5% control limit for the stage 1 model is shown, and NNU identified as possibly unusually low are highlighted**.**

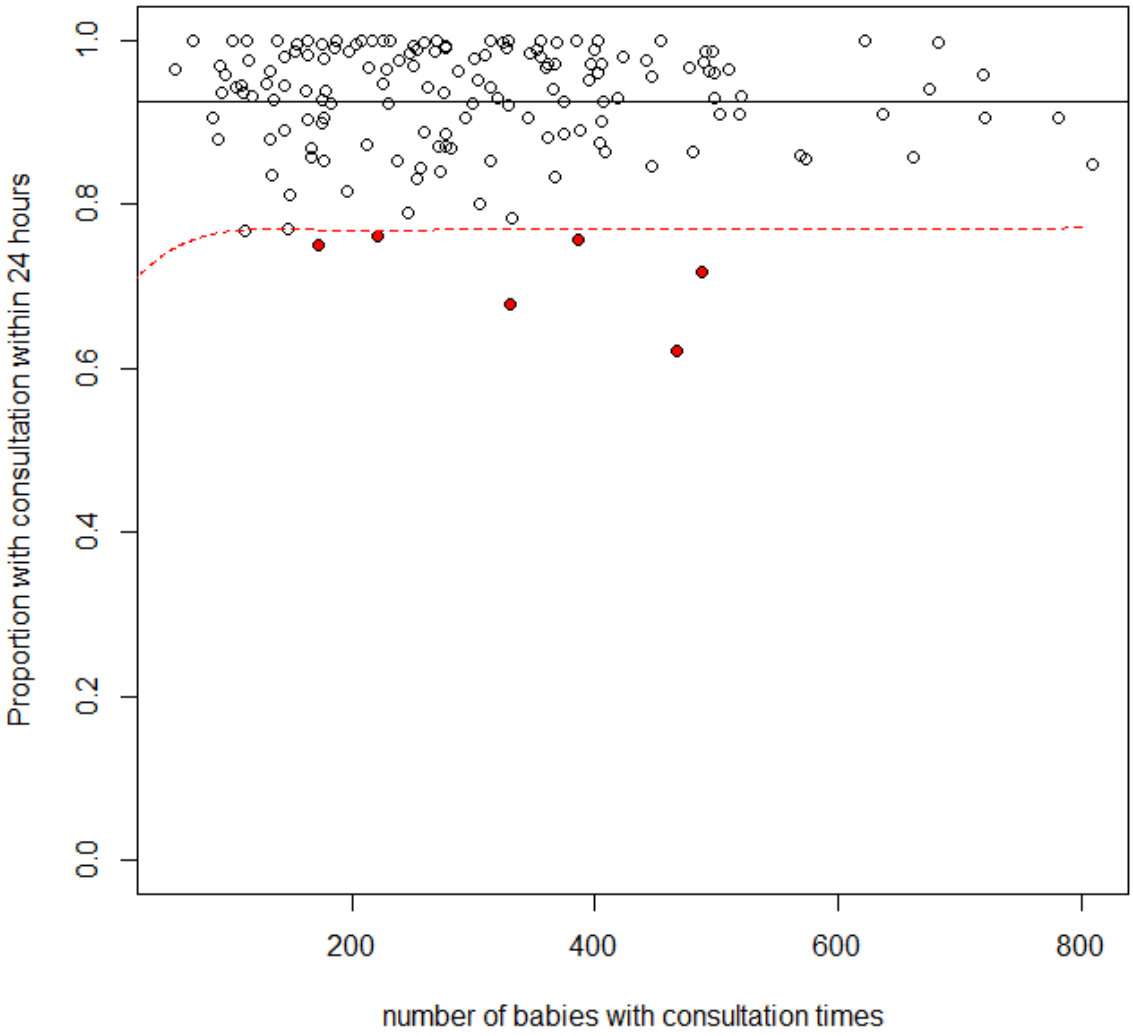
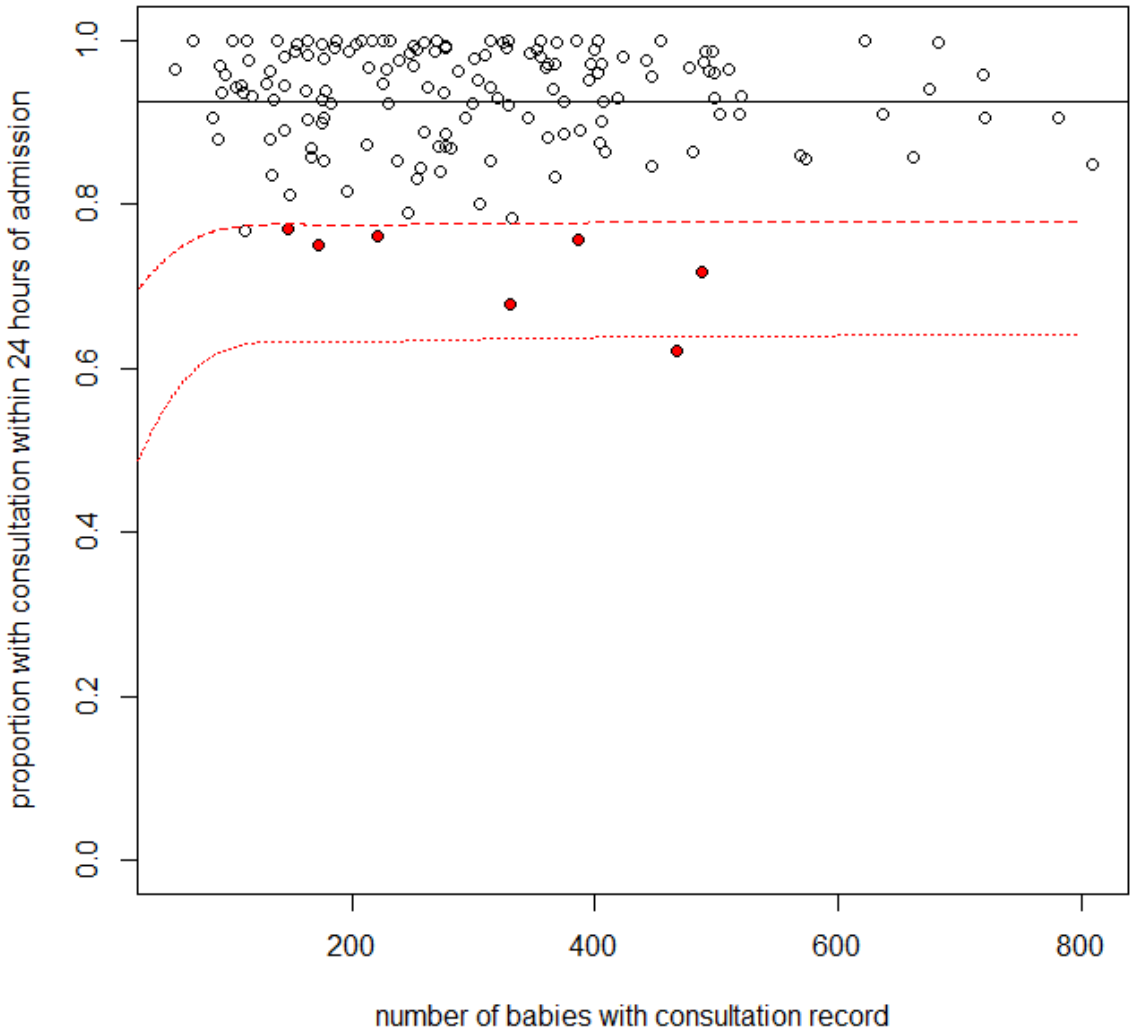


Figure 3: Funnel plot for proportion of babies with initial consultation time records who had a consultation within 24 hours of birth. The 2.5% and 0.1% control limits for the stage 2 model are shown, and NNU identified as possibly unusually low are highlighted.



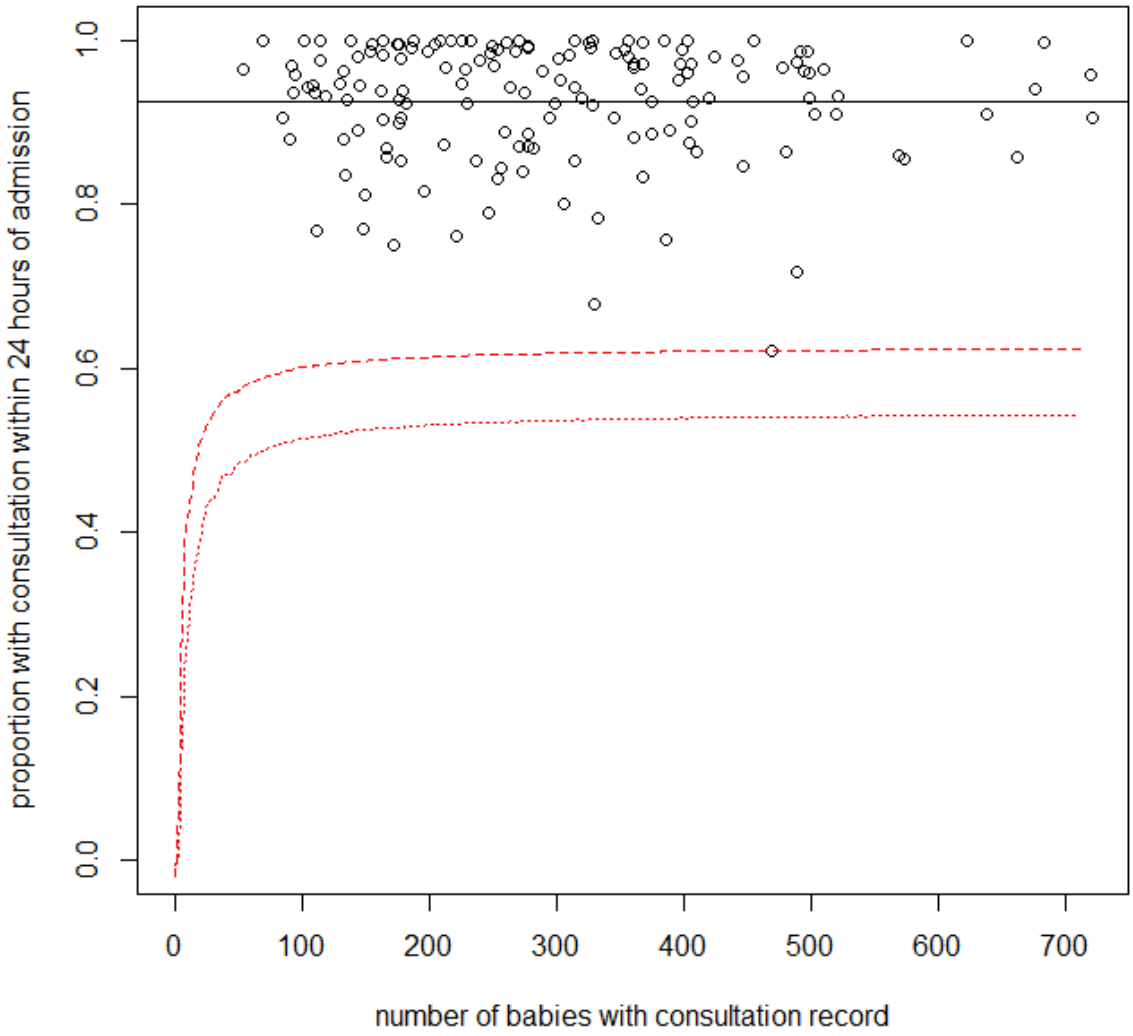
Adjusting for multiple testing:

Confirmation of 2014 low-outliers

On constraining the expected proportion of false positives to 10%, we found that no NNU would be identified as low outliers.

However one NNU was very close to the threshold, as shown in Figure 4. This NNU was identified as an outlier in 2013, with a large number of late consultations, and the percentage of babies with consultations on time has only improved marginally, from 60% to 62%. Further non-statistical investigation may be useful.

Figure 4: Funnel plot for proportion of babies with initial consultation time records who had a consultation within 24 hours of birth. The thresholds corresponding to the expected percentage of false positives 10% and 1% are shown.



Checking missing consultation times in participating NNU: we did not find any participating NNU would fall below the 10% false discovery threshold under the assumption that all missing values indicated failure to achieve the NNAP standard for consultation on time (worst case scenario). As the proportion of missing values is constrained to be low (<10%) among participating NNU, we would not expect to see many changes in outlier status under this assumption.

Comparison of non-participating NNU with thresholds:

Table 1 summarises available information for the non-participating NNU (i.e. NNU with more than 10% missing first consultation records). This table shows the proportion of babies with whose parents had a consultation within one hour of admission under the “best case” scenario that the parents of all the babies with missing consultation records did have a

consultation on time, and the “worst case” scenario that the parents of all these babies did not have a consultation on time. Comparing the best case and worst case proportions with the 10% false discovery threshold to identify unusual performance, we find that in the worst case several NNU (**The Royal Free Hospital, Great Western Hospital, Stepping Hill Hospital, Glan Clwyd Hospital, Glangwili General Hospital**) would be flagged as outliers. Under the best case scenario none of these would be flagged as outliers. In view of the large amount of missing consultation records, further non-statistical investigation of these NNU may shed further light on possible problems in consultations with parents.

3. **High performing units***:* There were 14 NNU with no missing initial consultation times and where there was an initial consultation within 24 hours for all babies – these are listed in Table 4. This is an improvement on the outcome in 2013, when seven NNU achieved this standard.

Table 4: NNU with all eligible babies having a recorded initial consultation time, and all initial consultation times being within 24 hours.

Unit Name	Eligible babies
Basingstoke & North Hampshire Hospital	217
Colchester General Hospital	314
Countess Of Chester Hospital	328
Epsom General Hospital	114
Queen Alexandra Hospital	454
Royal Berkshire Hospital	385
Royal Hampshire County Hospital	270
Salisbury District Hospital	163
St Helier Hospital	232
St Mary's Hospital, Isle of Wight	101
Stoke Mandeville Hospital	356
Sunderland Royal Hospital	208
University College Hospital	622
Wexham Park Hospital	403

References:

1. Ohlssen DI, Sharples LD, Spiegelhlater DJ. (2007) A hierarchical modelling framework for identifying unusual performance in health care providers. Journal of the Royal Statistical Society A, 170: 865-890.