National Neonatal Audit Programme (NNAP) Methodology and Statistical Analysis Plan

Published January 2025

V2.0 Published by RCPCH January 2025.

The Royal College of Paediatrics and Child Health is a registered charity in England and Wales (1057744) and in Scotland (SCO38299)

Cite as: National Neonatal Audit Programme Methodology and Statistical Analysis Plan v2.0, National Neonatal Audit Programme, January 2025. RCPCH: London.

Contents

Introduction

The purpose of this document is to provide a detailed methodological overview of the data and measure development steps and analyses contained in the National Neonatal Audit Programme (NNAP) annual reports.

The document is not designed to relate specifically to any one report year but will be periodically updated annually to ensure that it is in line with the latest NNAP methodology.

The document does not contain measure-specific methodologies (e.g. inclusion criteria, attribution, derivation tables), as these are covered in detail in the year-specific measures guides, which can be found at https://www.rcpch.ac.uk/work-we-do/quality-improvement-patient-safety/national-neonatal-audit-programme-nnap/measures.

The NNAP Annual Report contains key messages and recommendations by theme, case studies, support and resources for healthcare improvement, information for parents and carers, and the future direction of the Programme. The latest version is available to download at: www.rcpch.ac.uk/nnap.

Full year-specific results at unit level, interactive reporting tools and unit posters are available on NNAP Online at: https://nnap.rcpch.ac.uk/.

From 2024, the NNAP also publishes monthly updated results in the public domain via the frequent reporting dashboard, available at: https://www.rcpch.ac.uk/resources/nnap-data-dashboard

1. Data flow

The RCPCH receives the data contained in the NNAP dataset directly from System C, provider of the BadgerNet clinical system. The NNAP team processes data that is solely required for care quality and service improvement in relation to the aims and scope of the NNAP.

Figure 1 - Data flow describes this data flow and the feedback loop which disseminates results and recommendations to neonatal units, networks, and the wider system to inform and promote quality improvement.

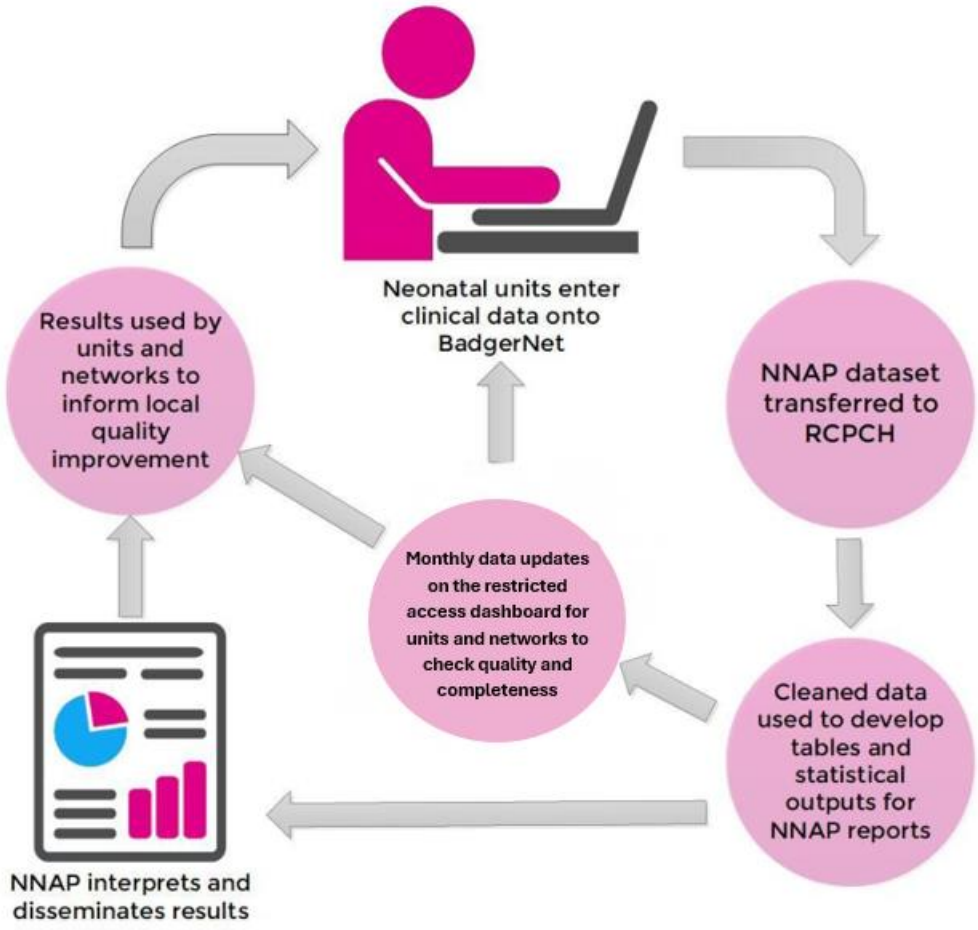

Figure 1 - Data flow

A full data flow diagram detailing data flows, data controllership, legal bases, storage, processing and outputs is available at: https://www.rcpch.ac.uk/nnap-data-flow-methodology

a. Live database linkage

Neonatal unit staff enter data into the BadgerNet platform (or systems that can interface with the BadgerNet database) as part of routine care whilst a baby is present on the neonatal unit and this data is stored on the BadgerNet database by System C.

Agreed data items contained within the NNAP dataset are synchronised to an NNAP database hosted by the RCPCH. This is done via SQL synchronisation between the Clevermed Azure NNAP Database patient data source and the RCPCH Azure database server over a TLS1.2 encrypted SQL<-> Synchronisation platform. This is created and operated by Microsoft on the Azure platform and is designed for secure data transmission and storage requirements.

As of 2022, the MESH mechanism to remove data about patients who have opted out at a national level via the National data opt-out (England only) is no longer applied before the dataset is transferred to the RCPCH, although patients can still directly opt out of inclusion in the NNAP.

b. Static database copies

Before any analysis is conducted on the data contained in the Live database hosted on the RCPCH Azure environment, static, pseudonymised versions of the database need to be copied on the Azure environment. This process allows a snapshot of the database to be taken at an agreed time, ensuring that analysis conducted on the data is replicable as the underlying data in that static version of the database is not updated and therefore does not change. At this point the database is also pseudonymised by the database administrator.

This static, pseudonymised copy of the database overwrites a single monthly report database each month, with an additional database copy saved for each annual report.

c. Pseudonymisation process

The RCPCH has approval to receive patient identifiable data without consent for England and Wales under Section 251 of the National Health Service Act 2006 however patient identifiable data is pseudonymised before analysis is conducted by the NNAP team to reduce the risk of data breaches.

When static database copies are made the database administrator uses SQL Server Management Studio to connect to the database and convert the identifiable and sensitive fields to pseudonymised versions of those fields. This is done by running a T-SQL script within SQL Server Management Studio which recreates the relevant data tables whilst implementing a combination of date transformations and hashing. The original

tables are then removed from the database. A full list of fields that are pseudonymised can be seen in Table 1 - Pseudonymised fields.

d. Data analysis

Once the static pseudonymised copy of the database has been prepared, this database is queried using the statistical software Stata, and analysis conducted on it. After each session the memory of the software is cleared, and no patient-level data is retained on the RCPCH local servers.

The NNAP dataset is defined in the NNAP data dictionary, available at: https://www.rcpch.ac.uk/work-we-do/quality-improvement-patient-safety/national-neonatal-audit-programme-nnap/measures

Table 1 - Pseudonymised fields

Table Name	Field Name	Field Description	Converted to
NNU2YearFollowup	AssessmentDate	the date and time of the follow up assessment	minutes from birth
NNU2YearFollowup	bailey_notes	free text field used in two year follow up questionnaire	generic text field
NNU2YearFollowup	date_death	the date and time of death	minutes from birth
NNU2YearFollowup	gastro_special_diet_text	free text field used in two year follow up questionnaire	generic text field
NNU2YearFollowup	griffiths_notes	free text field used in two year follow up questionnaire	generic text field
NNU2YearFollowup	growth_length_measuredHowText	free text field used in two year follow up questionnaire	generic text field
NNU2YearFollowup	growth_measurements_date	the date and time that growth measurements were taken	minutes from birth
NNU2YearFollowup	neurological_diagnosis_text	free text field used in two year follow up questionnaire	generic text field
NNU2YearFollowup	schedulegrowing_notes	free text field used in two year follow up questionnaire	generic text field
NNUCranialUSS	NoteTime	the date and time of the cranial ultrasound scan	minutes from birth, month, year
NNUDaySum	CareDate	the date of the care day	minutes from birth, month, year
NNUEpisodes	AdmitTemperatureTime	the date and time that the temperature was taken	minutes from birth, month, year
NNUEpisodes	AdmitTime	admission date and time of the episode	minutes from birth, month, year
NNUEpisodes	BirthDateMother	the date of birth for the mother	minutes from birth, month, year
NNUEpisodes	BirthTime	birth date and time of the baby	minutes from birth, month, year, month 44 weeks PMA, year 44 weeks PMA
NNUEpisodes	DateTimeOfDeath	the date and time of death	minutes from birth, month, year
NNUEpisodes	DischTime	discharge date of the episode	minutes from birth, month, year
NNUEpisodes	FirstConsultationWithParents	the date and time of the first consultation with parents	minutes from birth, month, year
NNUEpisodes	NationalID_CHI	the national ID for Scotland	MD5#
NNUEpisodes	NationalID_HNC	the national ID for Northern Ireland	MD5#
NNUEpisodes	NationalID_NHS	the national ID for England and Wales	MD5#
NNUEpisodes	NationalIDBaby	NHS/CHI number for the baby	MD5#
NNUEpisodes	NationalIDMother	NHS/CHI number for the mother	MD5#
NNUEpisodes	PostCodeMother	The postcode of the mother at time of birth	LSOA
NNUEpisodes	SteroidsLastDose	the date and time that the last steroids dose was given	minutes from birth, month, year
NNUEpisodes	SteroidsFirstDose	the date and time that the first steroids dose was given	minutes from birth, month, year
NNUROP	NoteTime	the date and time of the ROP screening	minutes from birth, month, year
NNUSepsisScreen	NoteTime	the date and time of the sepsis screening	minutes from birth, month, year
NNUUNICEF	firstexpressiontime	the date and time of first milk expression	minutes from birth, month, year
NNUUNICEF	firstrecievedexpressedmilk	the date and time the baby first received expressed milk	minutes from birth, month, year

2. Case ascertainment and unit participation

In usual practice, every baby admitted to a participating neonatal unit entered on the BadgerNet patient record system is eligible for inclusion in NNAP. The audit therefore achieves 100% case ascertainment in the participating organisations, unless a parent or carer has chosen to opt out of having their baby's information submitted to the audit. Babies receiving special care alongside their mother in transitional care areas or postnatal wards can also be entered onto BadgerNet, but it is known that some units do not enter data for such babies. For this reason, NNAP's measures do not concentrate on care outside neonatal units.

All neonatal units in England and Wales associated with a delivery unit are eligible to take part, including special care baby units (SCUs), local neonatal units (LNUs) and neonatal intensive care units (NICUs). As of April 1st 2022, the NNAP has permission from the Privacy and Public Benefit Panel for Health and Social Care (HSC-PBPP) to include neonatal units in Scotland. As of April 1st 2023, the NNAP also has permission to include neonatal units in the Isle of Man.

Where there is a change in unit name, unit level or network configuration, the NNAP will apply the status as at the end of the data reporting year. For example, if the configuration of a network changes on 1 April 2021, 2021 data will be presented as per the network configuration on 31 December 2021. Details on unit participation can be found in the appendix of the most recent audit report.

Name changes and closures of units, networks and trusts are managed both on an *ad hoc* basis (units and trusts can inform the audit of changes by emailing nnap@rcpch.ac.uk at any time), and through an annual review process in which emails are sent to all participating neonatal networks and units requesting notification of any changes to unit level, name, closures or other changes.

3. Data quality and completeness

Although the NNAP uses routinely collected data from the BadgerNet database, this data can be and is still subject to data quality and completeness issues that affect the quality and reliability of the audit findings. Data quality issues include:

1) Data fields that are optional on BadgerNet but are required for audit analysis not being routinely completed by units.
2) Data fields that are optional on BadgerNet but are required for audit analysis being omitted by staff members on an *ad hoc* basis.
3) Data fields that are mandatory on BadgerNet but are not completed by units that use direct interfaces with the BadgerNet database, and thus bypass the BadgerNet mandatory field requirements.
4) Database or BadgerNet errors that cause contradictory or missing results.
5) Missing data resulting from the database pseudonymisation process.

Issues 4) and 5) listed above are mitigated by data cleaning and validation checks on the database. Due to the high degree of control that RCPCH and Clevermed have over the processes related to these steps, such issues are typically resolved quickly and completely. They therefore rarely have a negative impact on data quality and completeness.

a. Restricted Access Dashboard

Issues 1), 2) and 3) are mitigated through the sharing of a Restricted Access Dashboard with units, which contain summary statistics for each measure and a list of pseudonymised patient identifiers corresponding to patients that are missing data or that may require review, for each measure.

Restricted Access Dashboard updates occur monthly, and units are expected to use the data contained in it to identify and resolve data quality and completeness issues generally (but not exclusively) related to missing data.

The Restricted Access Dashboard is hosted on Power BI, from within the RCPCH Microsoft tenancy, to which only the unit's clinical lead, the NNAP Data Analyst and Manager, and selected staff members at the unit, have access.

Units can update their data records to reduce missing data at any point over the year and their data will be updated in the following month. Units are given until 31st March of each year to ensure that their data for the prior year is correct and complete.

The pseudonymised static version of the database that is used for the generation of annual report figures is copied after 31 March to ensure all units have had an opportunity to clean their data.

Units can also access and review their data in real-time using the BadgerNet system reporting tools.

Unit feedback and communications

The Restricted Access Dashboard also provides assurance around the implementation of the measures' criteria on the data, with units feeding back to the audit wherever divergence between the measures guide and the attribution, inclusion or classification of patients is observed by units. Feedback provided by units is vital to ensure that analyses are correct and appropriate, and for the continuous improvement of the NNAP methodology and reporting.

Whenever new comments are raised by units, they are recorded in the NNAP query log, which is maintained by the NNAP Data Analyst and periodically reviewed. The NNAP team aims to respond to all queries within 3 working days, and where this is not possible holding emails are sent to inform the sender of the expected delay.

4. Data cleaning and validation

a. Validation

Every time a monthly or annual analysis is undertaken a new static copy of the live database is taken and pseudonymised for this purpose. Each time the monthly static database copy is overwritten it undergoes the following validation processes:

- The number of records in the episodic table is compared to the number contained in the last pseudonymised episodic table.
- Missing data percentages for each field in the episodic table are compared to the last pseudonymised episodic table.
- The total number of tables in the database is checked against the last pseudonymised database.
- Import and cleaning code is run on the new database to ensure that it runs without errors.

In addition to these validation checks, the following checks are conducted every time a new annual report cut of the live database is taken and pseudonymised for annual report analysis:

- The number of episodes removed by the data cleaning processes (below) is checked against the last pseudonymised database.
- Total rows for each table are checked against a list of expected total rows for each table provided by Clevermed Ltd.
- Checks are conducted between the episodic table and all other tables to ensure that all records in the ad hoc tables and the daily summary table have matching episodes in the episodic table.
- A review of the NNAP location lookup file is conducted to ensure that all expected units are presented in the data, and that all records match an expected provider unit.
- A draft run of the measure output proportions and N is compared to compiled results from the previous year, and large changes are reviewed.

b. Data cleaning

Once validated, a data cleaning process is applied to the database tables before creating the NNAP dataset used to produce the data included in this report. The pseudonymised database is cleaned using the following steps:

1. Episodes with no recorded birth year are removed from the analysis.
2. Episodes with no recorded gestational age are removed from the analysis.

3. Episodes with no recorded admission time are removed from the analysis.
4. Episode numbers with negative values are removed from the analysis.
5. Duplicated episodes are removed from the analysis, prioritising those with a greater number of completed fields.
6. Babies who have no episodes with days on a neonatal unit are removed from the analysis.
7. Babies with a gestational age of less than 22 weeks at birth are removed from the analysis.

Figure 2 - Episodes eligible for NNAP analysis describes episodes removed from the dataset during the data cleaning process for the 2021 data year.

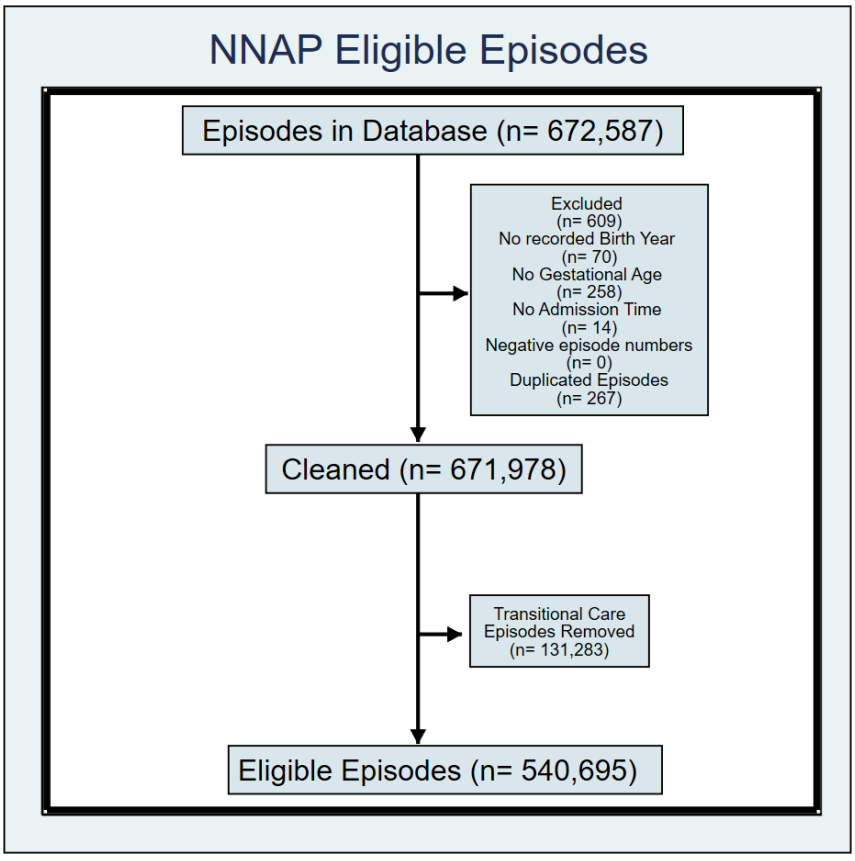

Figure 2 - Episodes eligible for NNAP analysis

Eligible episodes for each audit year include all neonatal admissions to NNAP units that were born between Jan 01 three years prior to the current data year, and 31 December of the current data year. For example, the 2021 data year used babies born between 01 January 2018 – 31 December 2021. It is important to note that not all babies in the eligible episodes are included in the measure analysis, which is dictated by measure-specific inclusion criteria. For details on inclusion criteria for each measure, please review the most recent version of the NNAP measures guide.

c. Location lookup matching

Alongside the data cleaning steps, it is necessary to match data from the pseudonymised database to NNAP unit names, codes, unit levels and networks. This is because:

- BadgerNet does not contain accurate data on neonatal unit levels (NICU, LNU, SCU).
- BadgerNet does not contain accurate data on which units are associated with which Operational Delivery Networks (ODNs).
- BadgerNet contains divergent names for units, with multiple NHS location codes and unit names being associated with one single unit.
- BadgerNet does not know which units are included in the NNAP and which are excluded.

For these reasons it is necessary to match the data in the database with NNAP unit names, unit codes, ODNs and unit levels. This is done by linking the episodic table of the database to a location lookup that converts episodes' unit of birth and provider unit to distinct NNAP unit codes and ODNs. Matching of these codes is made first on BadgerNet's NHS ODS code field, then for unmatched units that process is repeated on BadgerNet's unit name field. Units that cannot be matched are assigned to the OTHER category.

As the parameters of the live database, set by System C and RCPCH, only allow the sharing of episodes directly associated with NNAP units, all provider codes should match known NNAP unit codes. A place of birth code may relate to a different hospital than that of care provider, therefore not all place of birth codes match known NNAP units.

To reduce this number of unmatched place of birth codes in the dataset, babies whose place of birth is listed as Home or Transit have their network of birth updated to the provider network of their earliest episode.

5. Outlier identification and management

Outliers are identified with funnel plot analysis, using the national proportion as the comparison standard. The NNAP manages outlier status in line with its policy *Detection and management of outlier status*, available at: https://www.rcpch.ac.uk/nnap-data-flow-methodology#outlier-management-policy which is aligned with the HQIP guidance on the identification and management of outliers in England and Wales. All neonatal services identified as outliers for one or more NNAP measure are notified according to the policy prior to publication of the results.

Funnel plots for outlier detection are produced using a binomial formula. This linearly interpolates between anticipated counts of babies[1], to obtain consistently increasing or decreasing funnel curves. The funnel curves are then smoothed further for clarity, using local-linear regression via the Stata command 'npregress kernel'.

Overdispersion of the funnels via Spiegelhalter's formula[2] is also included. This can readily be applied to the binomial funnels and has been widely used in other clinical audits.

For the measures where differences in case mix have been accounted for by balancing, The standard error of the treatment effect is used to calculate outlier limits. Please see *Appendix 1 detailed description of the balancing process* for more information.

Upon completion of the funnel plot analysis, outliers are attributed to one of 5 categories, which correspond to:

1) If value <Lower 99.8% control limit then Alarm; or else
2) If value <Lower 95% control limit then Alert; or else
3) If value between Lower and Upper 97.5% control limits then Within expected range; or else
4) If value <=Upper 95% control limit then Excellent; or else
5) If value >Upper 99.8% control limit then Outstanding

The measures on which the full outlier management process is implemented are selected by the NNAP Methodology & Dataset Group with sign off by the NNAP Project Board; these can vary from year to year.

6. Public data sharing

The results of the NNAP are shared with the public through NNAP online, the Annual report and appendices, and through the NNAP Data Dashboard.

Details of the **annual report and appendices** are contained online and can be found here: https://www.rcpch.ac.uk/work-we-do/clinical-audits/nnap

The results of annual measure calculations and the associated statistical analyses are uploaded to the online portal **NNAP Online**. NNAP Online contains summary statistics for all units across all measures, as well as plots and charts related to the audit. The site is managed by Net Solving Ltd. and publishes the data on the same day that the report is released.

Stata is used to load, combine and reformat all annual report outputs into a .xlsx file that contains combined results for all measures into a format that can be used by Net Solving. This data is prepared once the annual report outputs have been finalised but before publication. This file is uploaded by Net Solving to a restricted access test site, on which the results can be interrogated before being enabled on the live site on the date of publication of the report.

The data is also used to remake spine plots, caterpillar plots, funnel plots and bar charts in a format different to that of the annual report outputs.

Annual results for selected measures are used within reporting by the Care Quality Commission (CQC).

More frequent data reporting is available via the **NNAP Data Dashboard**, which shows time series charts for each of the 10 NNAP reporting metrics. These totals are displayed as rolling annual proportions and are updated monthly.

The dashboard is created in PowerBI and uses the most up to date data available, with all months since the last release of the Annual Report listed as provisional. The NNAP Data Dashboard and more information about it can be found here: https://www.rcpch.ac.uk/resources/nnap-data-dashboard

7. Managing small numbers in the NNAP

The NNAP considers the risk of disclosure on a measure-by-measure basis from a variety of methods resulting from the publication of results from small numbers of cases. Given the frequent occurrence of small numbers in unit level, annualised reporting, applying blanket masking to all cells would significantly reduce the utility of published NNAP results for improvement purposes. To minimise the risk of disclosure, the NNAP does not publish demographic data about the cohort of babies included in the audit, which would have the potential to be used alongside published data for the audit measures to aid identification of a patient. In addition to this, small number suppression is applied to measures considered to be sensitive; reporting that a clinical activity took place, such as consultation with parents within 24 hours, is not considered to be sensitive.

Table 2 - Measures subject to small number suppression shows a table of measures to which small number suppression is applied in the final results.

Table 2 - Measures subject to small number suppression

NNAP measure/data item	Application of masking to unit level data.
Parents on ward rounds	Mask sub-measure data when main measure data is <2 or sub-measure proportion is >99% or <1%.
Bloodstream infection	Mask unit results where the denominator is <3 and the infection rate is not 0%.
BPD	Mask unit results where the denominator is <3 (for 3 year rolling results and individual year results) and the BPD or death rate is not 0%.
Necrotising enterocolitis	In outputs, combine the columns "Death prior to discharge with no NEC" and "no NEC". Combine the missing data columns "Alive at discharge" and "Died before discharge". Mask unit rates where the denominator is <3 and the NEC rate is not 0%.
Breastmilk on the first two days of life	Mask unit results where the rate of feeding with any mother's milk is either 0% or 100%, or where the denominator is <3.
Breastmilk feeding at discharge home	Mask unit results where the rate of feeding with any mother's milk is either 0% or 100%, or where the denominator is <3.
Breastmilk feeding at 14 days of life	Mask unit results where the rate of feeding with any mother's milk is either 0% or 100%, or where the denominator is <3.
Brain Injury	Mask unit results where the proportion of IVH 3/4, cPVL or PHVD is 100%, or where the denominator is <3.

8. Accounting for case-mix differences by balancing

Between 2018 and 2019, the statistical analysis process used by the audit to account for case mix differences between units and networks was matching. Due to complexities involved in the data flow change the report on the 2020 data used multilevel logistic regression to account for case mix differences – for more details on this decision and process used see Appendix A: Results and Methods from the 2020 report.

The report on 2021 data returns to a form of matching (called balancing) to account for case mix differences. In balancing, each unit and network is compared to the national dataset. For each comparison, the patients in the national dataset are weighted so that the national dataset matches as closely as possible the profile of matching variables of the unit or network being compared.

The balancing process is applied to the following measures: bronchopulmonary dysplasia, mortality, necrotising enterocolitis, and bloodstream infection. Although all measures have balancing applied to them, there are some differences in how it is implemented for each measure. These are described in detail in *Appendix 1 detailed description of the balancing process* and *Appendix 2 algorithm for balancing*.

a. Why the NNAP uses balancing over logistic regression

While many audits use logistic regression models to account for case mix differences between units, the NNAP uses balancing. Accounting for case mix using a balancing method is preferred to logistic regression because case mix adjustment with regression relies on:

- Models whose validity can never be established, only tested. That the validity of any logistic regression models used is not disproven does make them confirmed as valid.
- Assuming that the relationship between outcome and background characteristics is true in the whole population is valid for each unit or network analysed. In fact, it is plausible that the relationship between background characteristics and outcome differs between units.

- An assumption that units/networks' populations are homogenous, which they may well not be. Certain units may have more babies with unusual patterns of background characteristics, which will not be accounted for by a logistic regression model.

Balancing is analogous to a clinical trial, where the outcomes for a group of babies in a unit or network of interest are compared to a similar ('balanced) group of babies in the whole population.

Both methods are potentially limited by the availability of sufficient data to describe background characteristics of the patients.

9. Figures

a. Caterpillar plots

Caterpillar plots are generated for units and networks in all measures to help identify variations in the level of compliance between units, networks and unit levels.

In unbalanced caterpillar plots, each unit/network is represented by a vertical segment that extends from the lower to the upper confidence limit for the compliance proportion of the unit. These confidence limits are estimated as 2 standard errors above and below the unit/network proportion. The proportions (marked by discs) of compliance and estimated confidence limits are evaluated without shrinkage or imputation. Units are sorted according to their proportion for the given measure.

Catepillar plots are constructed in the same manner for balanced measures, except that units/networks are sorted in the ascending order of their treatment effect scores, and confidence intervals are obtained via bootstrapping.

From 2019, distinct colours are used to indicate the level of the unit. Figure 3 - Antenatal Steroids - Units (2022 data year) displays the unit-level caterpillar plot for antenatal steroids for the 2021 data year. The colours used are indicated in the legend (SCU — level 1, LNU — level 2, and NICU — level 3). For the networks, caterpillar plots are constructed similarly, except that the abbreviated names of the networks are printed at the horizontal axis.

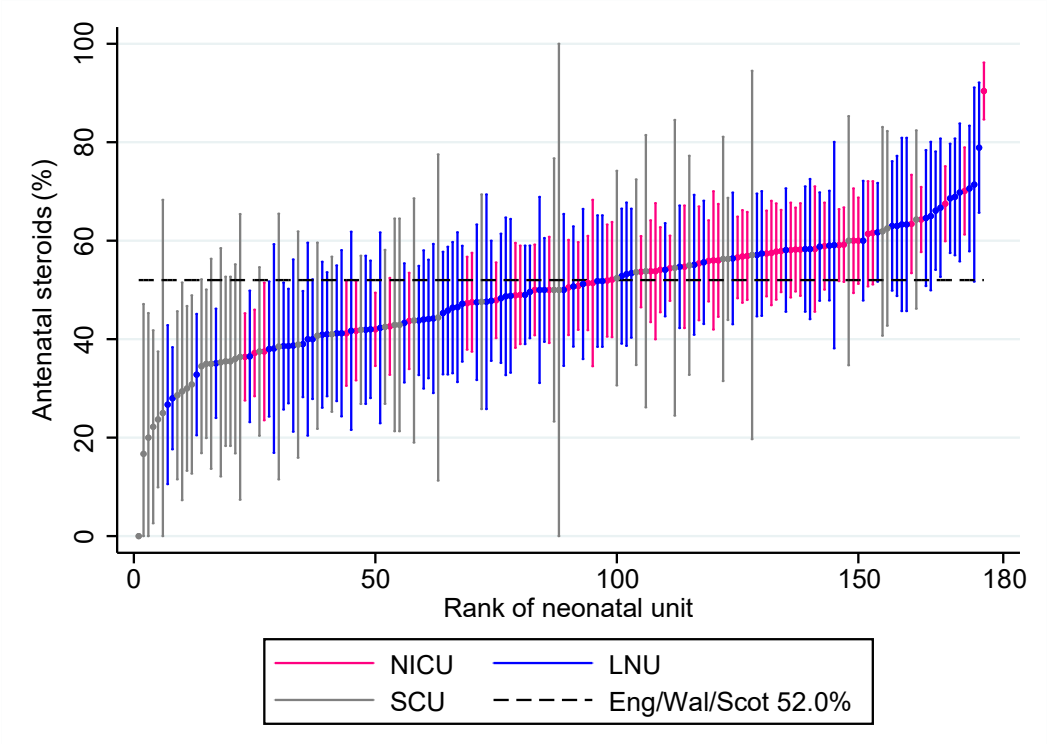

Figure 3 - Antenatal Steroids - Units (2022 data year)

b. Spine plots

A spine plot is a diagram in which the rates of several audit measures, for a unit or network, are presented in a single panel. It is a device to present the results of the audit in a compact form that facilitates comparisons of units, among themselves and with their networks, as well as networks with the whole domain. The segment extends from the lowest to the highest rate for the audit measure among the units in the diagram. Rates are scaled so that their network-level values are vertically aligned (marked by the dashes), and they fit the width available in the panel. The estimated rate for the unit is marked by a black disc, and the range in which the rate for the unit and the measure would be within the funnel limits is indicated by a grey strip. The network level rates are printed in a separate panel.

The black disc for a unit or network that is excellent or outstanding on a given audit item is located to the right of the grey strip and close to the right-hand limit of the segment, which corresponds to the highest rate among the plotted units or networks. The other extreme corresponds to poor performance. In a set of spine plots, we can identify units that have no weaknesses — all their discs are at the right-hand extremes; units that have no strengths — all their discs are at the left-hand extremes; units that have some strengths but also some weaknesses have their discs at either extreme of the segments.

Units that are average or mediocre for all items have their discs close to the corresponding network-level rates and within the grey strips.

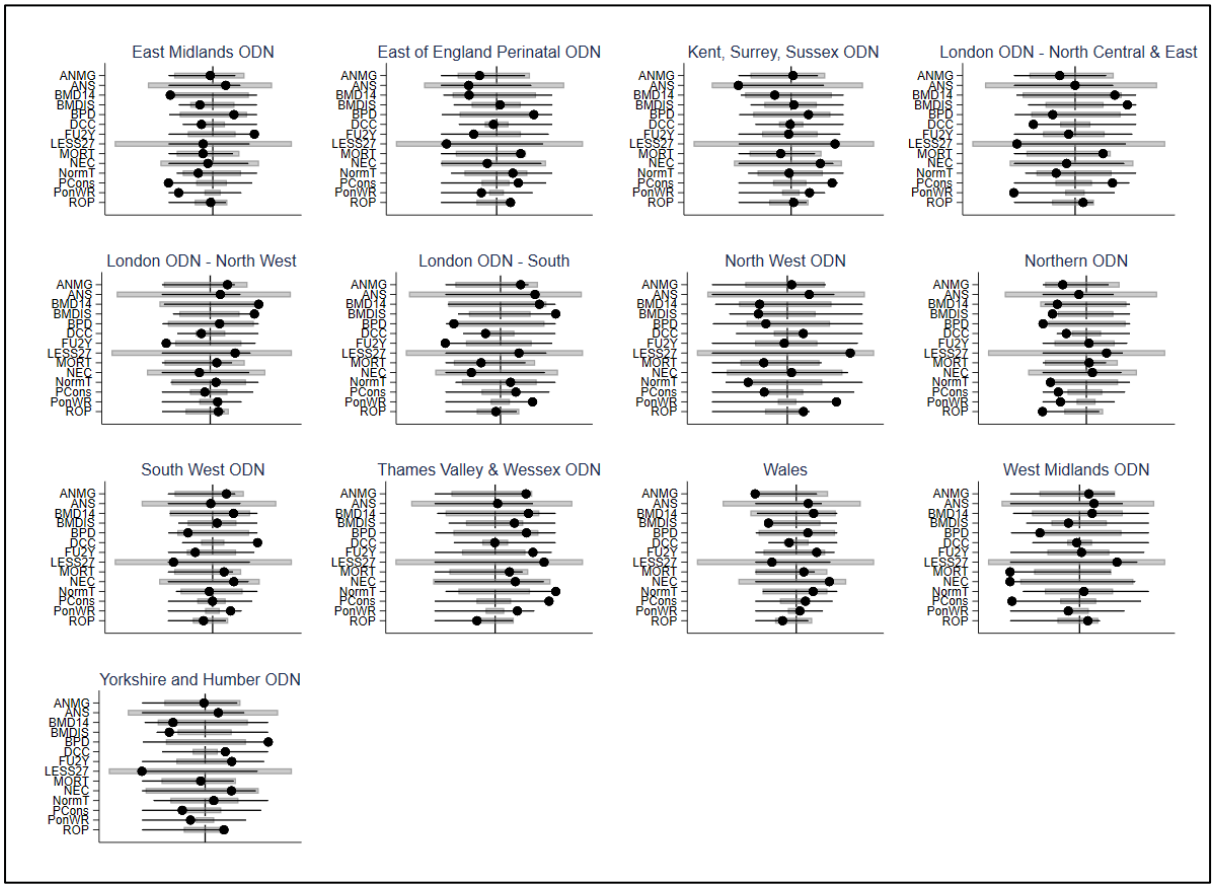

Figure 4 - Network spine plots (2020 data year)

c. Graphical summaries of breastfeeding

No single day on which breastfeeding status might be recorded can satisfactorily summarise a baby's breastfeeding over the entire stay in neonatal care. Moreover, no information about breastfeeding after discharge is available in the database. The rates of breastfeeding in units and networks on day 14 and on the day of discharge, presented in the standard tabular and graphical layouts, are supplemented, starting with the NNAP 2021 data year, by a graphical display in which breastfeeding throughout the stay, including the day of discharge, is represented.

The outcome variable is defined as a dichotomy for each day of stay in a neonatal unit. Breastfed on the day is coded as unity (1), and not breastfed as zero (0). Thus, a baby that stayed in care for 14 days has a breastfeeding record comprising 14 digits, each of them either 0 or 1. An example of such a record is 0011111 1100?00, with question mark '?' indicating missing entry (on day 12). For the definition of the population (the eligible

babies and attribution), see Sections 3 and 4. In particular, the data span three years according to the date of birth.

Breastmilk feeding in a unit is graphically described by a colour-plot. Figure 5 – Thames Valley and Wessex breastfeeding (data year 2022) presents a set of examples, for units in Yorkshire and Humber Neonatal ODN. The axes of the plot are postnatal day (horizontal) and number of babies (vertical). An (imaginary) vertical line drawn on any given postnatal day comprises up to six segments of distinct colours, listed from bottom to top:

- Black at the bottom accounts for babies not fed at all (nil by mouth).
- Grey colour is for babies who were in a neonatal unit and were fed enterally but not breastfed on the day.
- Green colour is for babies who were breastfed on the day.
- Light green colour is for babies who have already been discharged and were breastfed on the day of discharge.
- Light grey represents babies who have already been discharged and were not breastfed on the day of discharge.
- Off-white colour at the top represents babies whose breastfeeding status is not known.

Plots are drawn for the first 90 postnatal days. They are drawn only for units with at least 70 babies in the data. If the entry for the day of discharge is missing, it is replaced by the breastfeeding status on the immediately preceding day. If the status is missing on both days, 'not breastfed' is imputed, unless the status is missing for many babies. Grey colour is reserved for missing breastfeeding status at discharge. Sub-sequences 0?0 and 1?1 are replaced by 000 and 111 respectively. That is, if breastfeeding status is missing for an isolated day and the states on the immediately preceding and following days coincide, then this status is imputed for the day with the missing record.
The colour-plots can be interpreted only by comparing the sizes of the areas (and their shapes) for the units (within a network) or of a unit with its network. It is therefore essential to present the colour-plots in sets of a selected network on a single page.

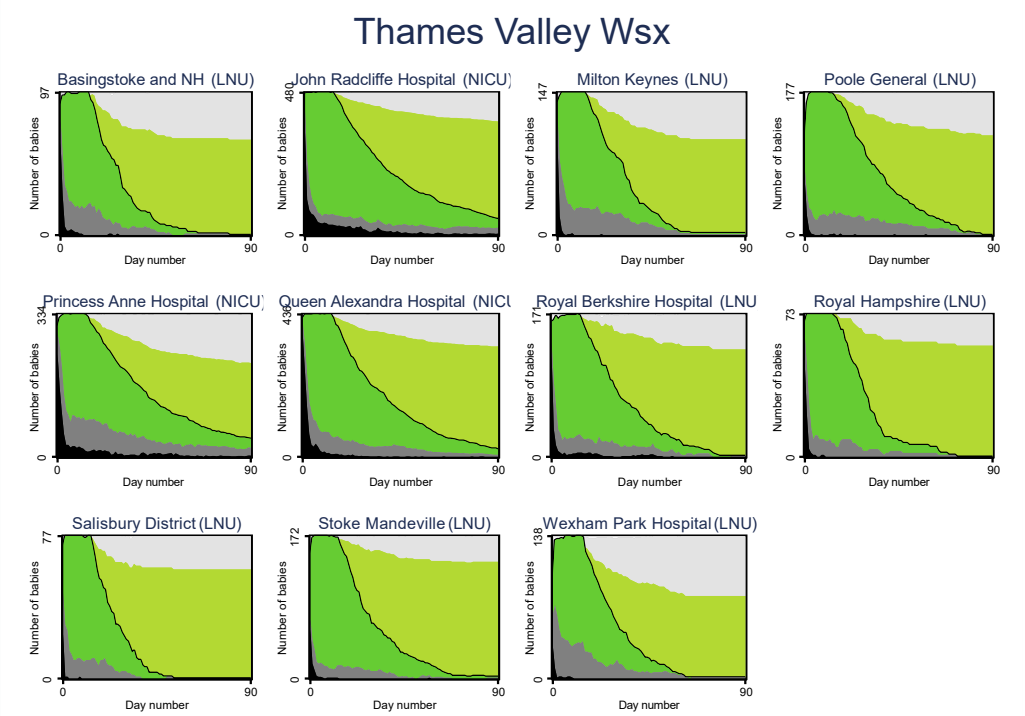

Figure 5 – Thames Valley and Wessex breastfeeding (data year 2022)

d. Nurse staffing diagrams

Nurse staffing plots designed to highlight annual shortfalls were added to the outputs for the 2021 data year. A unit's workloads, shortfalls and excesses for a year, or another period, are presented in a plot of the workloads and nurses available over time. The workloads are drawn by a solid black line. It has to be smoothed to avoid sudden changes and to better represent the overall trend. Orange patches under the workload curve represent shifts with shortfalls and light blue patched above the workload curve are for excesses. These patches also involve smoothing, so that periods covered by the orange colour may contain some shifts with excess of nurses, but shortfalls dominate. The year represented on the horizontal axis is split into months, indicated by vertical dashes.

Smoothing is implemented using the normal kernel with standard deviation set to 7, selected by trial and error. In this method, the value of the smoothed curve at any particular point (shift No.), referred to as the focal point, is calculated as the weighted average of the values (of the workload or the number of nurses present) for all the shifts, with weights equal to the density of the normal distribution (the normal curve) centred around the focal point, with standard deviation 7. Blue patches for shifts with (dominance of) excess of nurses are of no interest to the audit, but may be useful for the unit in analysing past scheduling and identifying where improvement could be made.

The dots at the bottom of the plot represent individual shifts, with the blue dots indicating sufficiently staffed shifts and orange dots indicating shifts with staffing shortfalls.

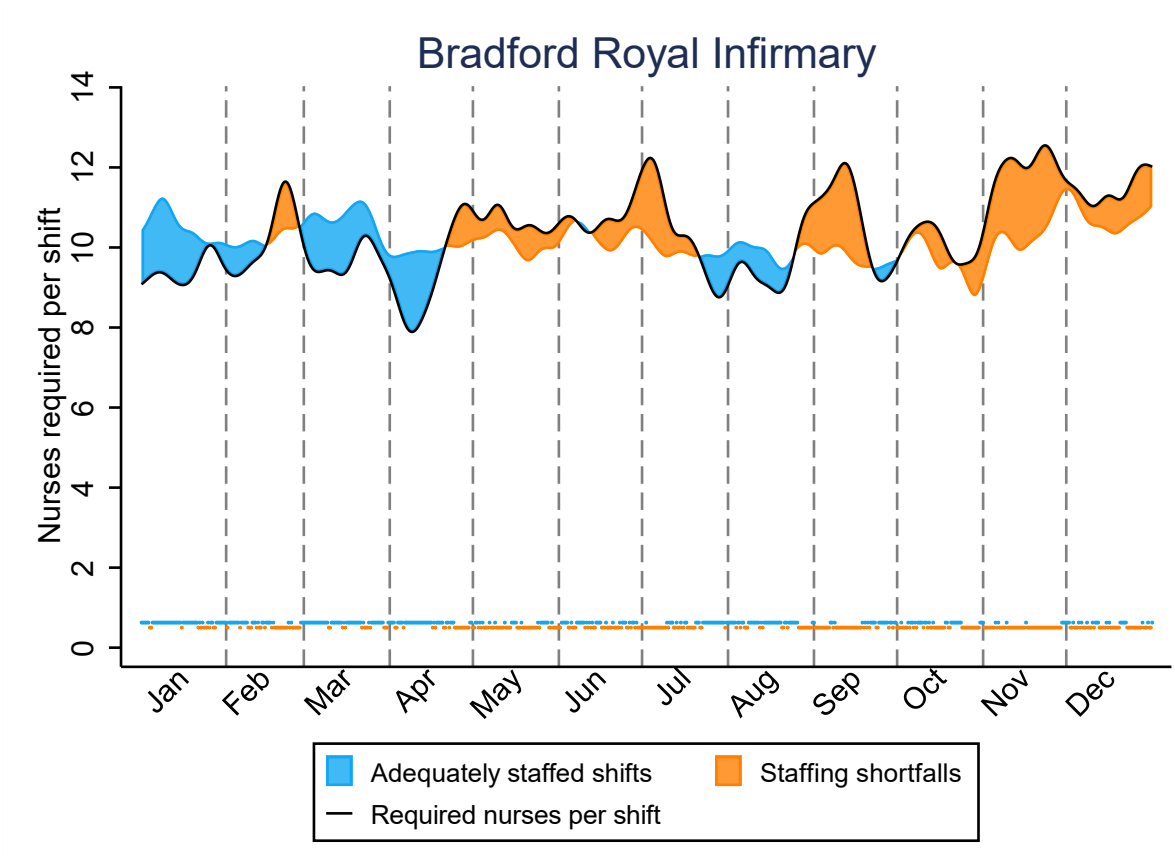

Figure 6 Bradford nurse staffing plots (2022 data year)

b. Other figures

In addition to the figures listed above, the audit regularly contains additional graphs and plots related to specifics in any given year. These include but are not limited to bar charts, longitudinal stick plots and scatter plots of measure proportions by unit.

Such figures are produced on an *ad hoc* basis with the goal of focusing attention on areas of interest pertaining to a specific audit year.

Funnel plots are described in Section 5.

Appendix 1 detailed description of the balancing process

a. Bronchopulmonary dysplasia

Accounting for case-mix differences by balancing

In the audit, bronchopulmonary dysplasia (BPD), is defined as requiring breathing support at 36 weeks of gestational age (GA). A baby that died before reaching this age is also included in the numerator, therefore the measure is titled `significant BPD or death'. The analysis is based on data for the past three years (e.g., 2019–21 in the 2021 Audit). It includes babies born before reaching 32 weeks GA; that is, GA < 32^{+0}.

The rate of BPD or death estimated by these rules cannot be compared straightforwardly with its counterpart in the 2016 Audit (based on data for years 2014–2016), because cases of death were not counted earlier. Also, the rates are calculated using different rules for imputation in the two Audits; no imputation has been used since the 2017 Audit. Since then, the rate of missing values has declined to such a level that this issue can now be regarded as moot.

For the analysis of a network, referred to as the *focal network*, balancing is used to find a set of weights for the babies from the entire country, for which the weighted means for continuous background variables and weighted proportions for categorical background variables are as close to the sample (unweighted) means and proportions of these variables in the focal network as can be arranged. These weights are referred to as the *balancing weights*.

Balancing is an adaptation of matching, which was used in NNAP for BPD between 2016 and 2019. Matching can be described as balancing in which all the weights are equal to unity (presence in a match; inclusion) or zero (exclusion). A separate analysis is conducted for each network. The analysis is based on Zubizarreta (2015)[3], with an adaptation that converts the problem to finding the minimum of a quadratic function (see Appendix 2 algorithm for balancing).

The *sample rate* of BPD for a network is defined as the percentage of babies included in the Audit from the network who were assessed to have BPD or death at 36 weeks of postmenstrual age. The *matched rate* of BPD for this network is defined as the weighted percentage of babies in the entire (national) audit for BPD. The balancing weights are used to evaluate the matched rate.

The matched rate of BPD is interpreted as the would-be rate of BPD of the babies from the network *if* they were dispersed for treatment throughout the domain of NNAP babies. The *treatment effect* is defined as the difference between the sample rate and the matched rate for the network. Negative values indicate a performance superior to the domain – as a group, the network's babies fared better in the network than if they were dispersed throughout the domain. Positive values indicate a performance inferior to the domain – an excess of cases over what one might expect.

The treatment effect is an estimate. Its standard error is evaluated by reference to binomial distribution. This facilitates a standard outlier analysis, as described in Chapter 5, and the drawing of caterpillar and funnel plots.

The method used, in which we compare a network with the domain by reference to the babies from the network, is called *indirect standardisation*. It compares the domain to the network, on terms of the network, and so it is useful for comparing the network to the country, but less so for comparing two networks. In direct standardisation, we define a *template*, or a synthetic reference group, by specifying a group of babies by their background profiles and compare the would-be performances of the networks on this group. That is, each network is assigned the same (clinical) task, so the comparison is fair, although unevenly relevant because the backgrounds of the babies in the reference set are more typical in some networks than in others.

Priorities and imbalances

In the algorithm for balancing, a set of parameters, called *priorities*, have to be specified. They reflect the relative urgency of matching the weighted and unweighted means (or proportions). Altogether, 19 parameters must be set, including one for the intercept. The role of the intercept is to ensure that the total of the weights is close to unity. The parameters are set to 10.0 for the intercept and gestational age, 3.0 for birthweight, and 0.25 for the remaining 16 parameters.

The parameters are relative to the urgency to reduce the dispersion of the weights to minimum, which is related to the efficiency of the resulting estimate of the treatment effect.

The imbalance for a background variable is defined as the scaled (standardised) difference of the weighted mean of a background variable for the domain and the unweighted mean of the same variable for the focal network. These imbalances are the principal diagnostics of the method. The analyst should check that they are all sufficiently small (close to zero); their sign is immaterial. All the imbalances should be smaller than 0.01 in absolute value and the mean of their absolute values should be smaller than 0.005.

One or several priorities may (and should) be increased when an (absolute) imbalance stands out or the weights add up to a value different from 1.0.

Table 3 - List of matching variables

Variable Name	Description	Type
ethnicitybaby	The ethnicity of the baby	Binary
ethnicitymother	The ethnicity of the mother	Binary
fetusnumber	If the baby is a multiple	Binary
labouronset	The type of labour	Binary
problemspregnancymother	Free text field of pregnancy problems	Binary
sex	The sex of the baby	Binary
smokingmother	If the mother smoked during pregnancy	Binary
agemother	Age of mother on date of baby's birth	Ordinal
birthtime_month	The month in which the baby was born	Ordinal
birthweight	The weight of the baby in grams	Ordinal
birthyear	The year in which the baby was born	Ordinal
gestationweeks	The gestational age of the baby in weeks	Ordinal
mother_lsoa	LSOA of the mother's address, linked to deprivation quintile	Ordinal
previouspregnanciesnumber	Number of previous pregnancies related to mother	Ordinal

The list of background variables was revised in February 2020, to be applied in the analysis of the 2020 Audit. At the same time, the decision was made to apply the same method, indirect standardisation by propensity matching, also to Mortality; see Section *Mortality*. Since 2021, the method has also been applied also to Bloodstream infection and Necrotising enterocolitis.

Caterpillar plots are drawn for the sample rates and for the treatment effects on a single page.

A funnel plot is drawn for the treatment effects; the horizontal axis is for the standard error and the vertical axis for the treatment effect. The thresholds that divide usual (satisfactory) networks from networks labelled as 'Alert' and 'Excellent' are straight lines with zero intercept and slopes +2 and –2. The thresholds for 'Alarm' and 'Outstanding' outliers are straight lines with slopes +3 and –3.

Units. Distance-based matching

The method used in the analysis of networks (BPD and Mortality) has been adapted to the analysis of units and implemented in a suite of R functions. This adaptation follows the same principles as the discarded method that used distance-based matching. Instead of matching, balancing is used. Balancing is a natural development from

matching. Balancing uses weights defined in the population of babies (a weight for each baby). Matching can be described as a special case of balancing, with each weight constrained to be equal to either zero (0) or unity (1), corresponding to being involved in no match (0) and being involved in a match.

A separate analysis is conducted for every unit at levels 2 and 3 (from 2022, analysis is also conducted on level 1 units), just as it is for every network in the analysis of networks. In the analysis of a unit, weights are found for the babies in the entire population so that the weighted mean (or weighted proportion for a binary variable) of every background variable in the population is as close to the ordinary (unweighted) mean or proportion of the same variable in the unit as can be arranged. The background variables 'compete' for proximity of the means in the unit and the population – improvement for one variable may be attained at the cost of spoiling the good balance for another. This competition is moderated by the tuning parameters called *importances* (or priorities). Their values are set higher for the variables *prima facie* strongly related to the outcomes (gestational age and birthweight), for which a tight balance is more important. The importances are set by trial and error; the principal criterion for them is that all the scaled differences of the weighted means in the population and the corresponding unweighted means in the units should be in as narrow ranges as possible. As a guideline, the largest imbalance should not exceed 0.001 in absolute value. Being one (population) standard deviation apart corresponds to the imbalance of 1.0.

The background variables comprise the same set as for the analysis of networks. Of course, they may be revised in the future; this would require only minor changes in the code. There is no upper limit on the number of background variables that can be included. In general, more variables and greater variety of them make the analysis more credible, but it is essential that they be background variables, unaffected by the assignment to the unit or the population in the hypothetical experiment that motivates the analysis.

Computational details

The weighted mean of a variable, as well as the imbalance for this variable, is a linear function of the weights, and its square is a quadratic function. The objective function that is minimised is the sum of squared deviations weighted by the importances. This function is quadratic, and so its minimum is (relatively) easy to find and the search involves no iterations. In fact, the solution is $\mathbf{w} = \mathbf{H}^{-1}\mathbf{s}$, where **H** is a square positive definite matrix and **s** is a vector. **H** is $N \times N$, where N is the number of babies in the population. Inverting this matrix numerically is not feasible. A recursive algorithm for its inversion is applied. It exploits the form of **H** as a matrix that is equal to the identity

matrix plus a matrix of relatively small rank equal to the number of background variables plus one. Neither **H** nor $\mathbf{H}^{-1}$ have to be formed in the process. The algorithm is computationally feasible for many more babies in the population (tens of thousands) and many more background variables (50+).

Alternatives

For measures with rare outcomes (BSI and NEC) we apply balancing (weighting) on gestational age (weeks) within three categories of units defined by level and status: level 2 (LNU), surgical units (they are all NICUs) and non-surgical NICUs.

The unit-level sample rates and estimated treatment effects and their confidence intervals are presented in caterpillar plots using the same layout as for networks, except for omitting the names of the units.

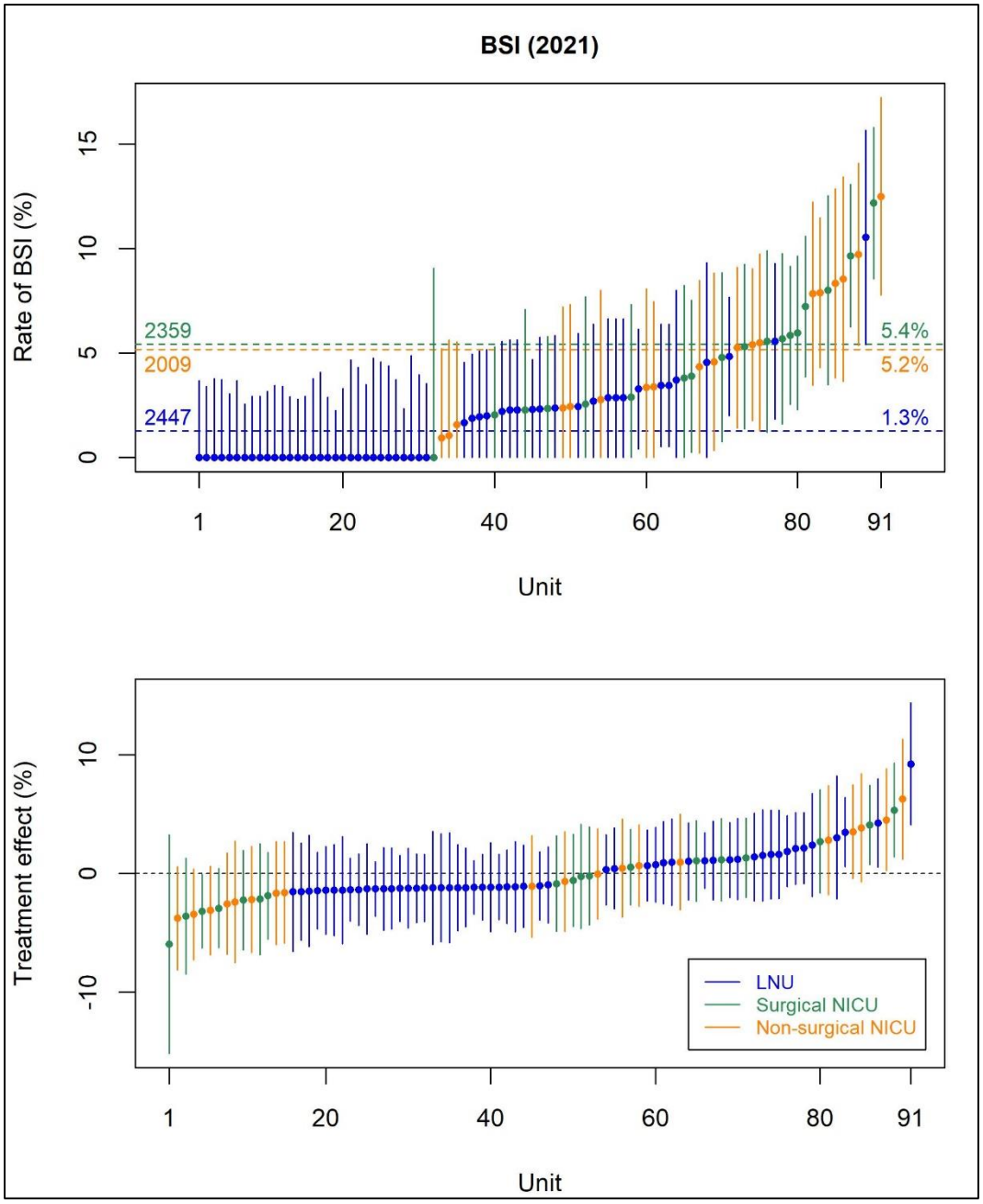

Figure 7 - BSI cat plots (2021)

Separate figures can be drawn for levels and states of the units. Throughout the levels and states are indicated by colour:

level-2: blue; level-3 surgical: green; level-3 non-surgical: orange.

The rates of BPD within these categories are indicated by horizontal dashes accompanied at the left-hand margin by the number of babies.

Funnel plots are presented for each level and status separately, using the same layout as for the networks. Outliers may be reported.

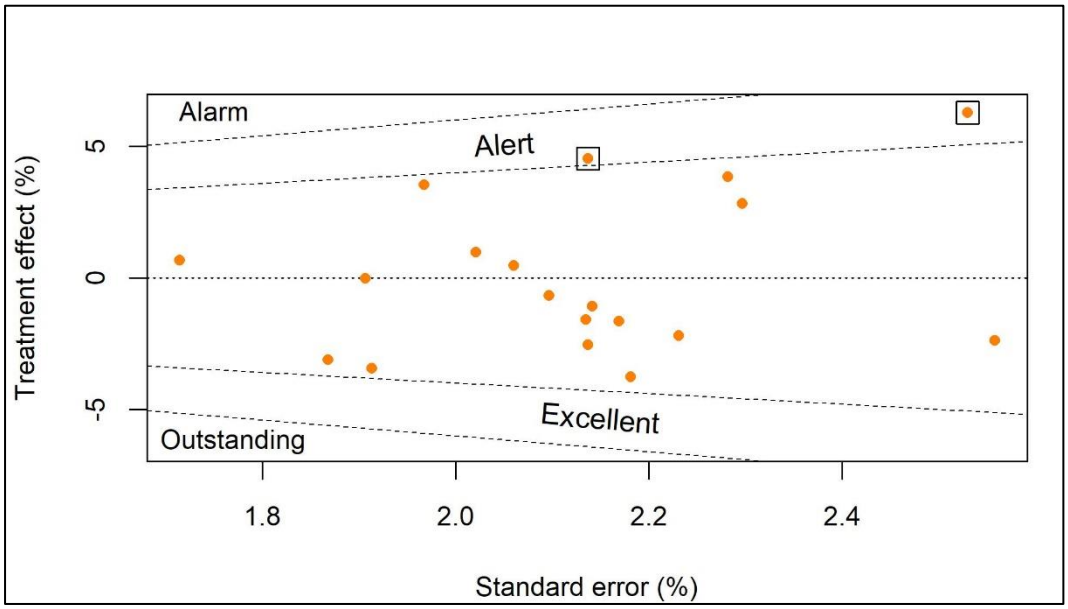

Figure 8 - BSI non-surgical NICU funnel plot (2021)

b. Mortality

As an Audit measure, Mortality is defined for a baby as having died before reaching 44 weeks of postmenstrual age. A baby discharged from a unit before reaching this age is regarded as having survived. The measure is defined for babies born at a gestational age (GA) of lower than 32 weeks. Qualified babies admitted to a participating unit in the preceding three-year period included in the analysis. An analysis is conducted also for the subset of babies born earlier than 28 weeks GA. Analyses are conducted only at the network level.

The method of balancing (indirect standardisation) is applied to adjust for the differences in the background profiles of the networks. The same set of background variables is used for balancing as for BPD; see Chapter 10. a. 'Bronchopulmonary dysplasia' for details, including the format and layout of the outputs and identification of outliers.

A separate analysis is performed for the subset of babies with GA at birth below 28 weeks. This group form about 30% of the babies, but accounts for around 70% of deaths. Mortality validation is not used in the analysis; data from all the units are used.

c. Necrotising enterocolitis

Necrotising enterocolitis (NEC), as an audit measure is defined by confirmed diagnosis, surgery, or post-mortem. Analyses are conducted at the network and unit level. For networks, adjustment for the babies' background profiles is applied by balancing; see Chapter 10. a. 'Bronchopulmonary dysplasia' for the rationale and outline of the method. For units, balancing is applied only with respect to gestational age and level of the unit. Only LNUs and NICUs (surgical and non-surgical) are analysed.

Networks

The sample rate of NEC for a network is defined as the percentage of the network's caseload that were ascertained as cases of NEC. After balancing, applied separately for each network, the matched rate is defined as the weighted percentage of the domain's caseload that was ascertained as a case of NEC. The treatment effect is defined as the difference between the sample rate and the matched rate of NEC.

The same background variables are used for balancing, and the same values of the priorities are set for every network as for the networks in the analysis of BPD. The same formats and layouts for tables and figures are used as for BPD.

Units

A separate analysis is conducted for each unit, and the units are grouped by their level (LNU or NICU) and surgical status. A unit in a group (e.g., surgical NICU) is balanced within the sub-domain on all the surgical NICUs. Within such a group of units, balancing is applied with respect to GA in completed weeks truncated at 24 weeks GA. That is, the categories used for balancing are 24, 25, ..., 31 weeks GA, with babies born at 23 weeks or earlier included in the category 24. The matched rate for a unit is obtained by combining the within-category rates with weights equal to the number of babies from the focal group in the same category.

This simple method of balancing is used because many units have only a few (and some no) cases of NEC. The same method is used for the unit-level analysis of Bloodstream infection for which the rates are of similar magnitude. For both audit measures, balancing on GA and type of unit is essential; the rates of both conditions decrease with GA and are higher in NICUs.

The principal result of the analysis is the estimated treatment effect and its standard error for each unit. For layout and format of tables and figures refer to Chapter 10. a. 'Bronchopulmonary dysplasia'.

d. Bloodstream infection

As an audit measure in NNAP, Bloodstream Infection is defined for babies born before reaching gestational age of 32 weeks. The unit can be attributed to a baby without any ambiguity only when the baby was not subjected to any transfers before their final discharge (or death). The network can be attributed to these babies as well as babies who underwent one or several transfers, but only when all the units involved were in the same network. Consequently, separate datasets are compiled for the unit- and network-level analyses. They contain the relevant caseload for the past year. The network-level dataset is more extensive.

Units are requested to validate the data they provide. Unit-level analysis, including outlier analysis, is conducted only for units with validated data. Network-level analysis is conducted for all the data, including data from units that failed to validate them. Unit-level analysis is restricted to LNUs and NICUs, and a distinction is made between surgical and non-surgical units. Babies from SCUs are included in network-level analysis. It is anticipated that there are very few such babies because every effort should be expended to ensure that a very preterm born baby is delivered in a ward adjacent to a NICU.

Analysis of units

Analysis of units uses balancing on GA and type of unit (LNU, non-surgical NICU and surgical NICU). The treatment effect for a unit is estimated by the difference of the sample rate and the matched rate, which is a combination of the population rates within the GA's (24 – 31 weeks, GA at birth truncated at 24 weeks) for the same type of unit. See 8. c. 'Necrotising enterocolitis' for details. Caterpillar and funnel plots are drawn and a list of outliers is compiled. For layout and format of tables and figures refer to section 8. a. 'bronchopulmonary dysplasia'.

Analysis of networks

Networks are analysed by balancing on the established set of background variables listed in **Table 3 - List of matching variables**. The treatment effect is estimated by the method described in Section 8. a. bronchopulmonary dysplasia. The output is an estimate and standard error of the treatment effect for each network, on which the caterpillar plot and the assessment of the outlier status are based. The standard format and layout are used for all tabular and graphical outputs.

e. Non-invasive respiratory support

Introduced to the NNAP in the 2022 data year, non-invasive respiratory support measures the number of babies who did not receive did not receive invasive respiratory support on

any of their first 7 days of life. As the proportion of babies receiving invasive respiratory support is known to decrease with gestational age, Non-invasive respiratory support is also subject to balancing on gestational age.

For units, this process is identical to that implemented for the NEC and BSI measures. For Networks this process is different in that is only balances of gestational age, whereas other balanced measures use a wider range of background variables for their Networks. For both units and networks, the outputs is an estimate and standard error of the treatment effect.

Appendix 2 algorithm for balancing

This appendix describes the algorithm for balancing, applied in the network-level analysis of BPD. It is based on Zubizarreta (2015) who proposed to minimise the dispersion of the weights, var(w), subject to upper limits on the imbalances, defined as the deviations of the means/proportions of the background variables in the focal network from the weighted means/proportions in the entire domain:

$$|\mathbf{x}_{Ak}'\mathbf{1}/n - \mathbf{x}_{Bk}'\mathbf{w}| < d_k$$

and $\mathbf{w}'\mathbf{1} = 1$, where d_k is the upper limit for the absolute imbalance, $\mathbf{x}_{Ak}$ is the vector of values of background variable k in the focal network and $\mathbf{x}_{Bk}$ is the corresponding vector in the domain, $k = 1, \ldots, K$; $\mathbf{1}$ is the vector of unities and n the caseload of the focal network. Note that, in general, $\mathbf{x}'\mathbf{1}/n$ is the mean of the vector $\mathbf{x}$ of length n. This problem of constrained optimisation is converted to one of unconstrained optimisation as minimising the weighted total of the squared imbalances $(\mathbf{x}_{Ak}'\mathbf{1}/n - \mathbf{x}_{Bk}'\mathbf{w})^2$, the squared deviation of the total weights $(\mathbf{w}'\mathbf{1} - 1)^2$ and the variance $(\mathbf{w} - \mathbf{w}'\mathbf{1}/n)^2$. The weights, to be set by the analyst, reflect the relative priorities of minimising each (squared) term.

This leads to a compact formulation of the problem, using matrix notation, as minimising the quadratic function $\mathbf{w}'\mathbf{H}\mathbf{w} - 2\mathbf{w}'\mathbf{s} + C$, where

$$\mathbf{H} = \mathbf{I} + \mathbf{X}_B'\mathbf{U}\mathbf{X}_B$$

$$\mathbf{s} = \mathbf{X}_B'\mathbf{U}\,\mathbf{x}_A$$

and C is a constant not important to what follows. Here $\mathbf{I}$ is the identity matrix, $\mathbf{X}_B$ is the matrix of the background variables for the domain, $\mathbf{x}_A = \mathbf{X}_B'\mathbf{1}/n$ the vector of the means of these variables in the focal network and $\mathbf{U}$ is the diagonal matrix with the priorities on the diagonal.

The solution is $\mathbf{w} = \mathbf{H}^{-1}\mathbf{s}$, and this expression can be evaluated without numerical inversion of any large matrices because $\mathbf{H} - \mathbf{I}$ is a matrix of relatively low rank, no greater than the number of background variables, K.

The solution $\mathbf{w}$ may have some negative elements. The corresponding babies are excluded and the algorithm repeated until the number of babies with negative weights is small, and so is their negative total. Negative weights can be interpreted as belonging to babies who are atypical in the focal network and not useful for balancing.

The principal diagnostics is based on the imbalances $\mathbf{x}_{Ak}'\mathbf{1}/n - \mathbf{x}_{Bk}'\mathbf{w}$ for variables k =1, ..., K and the deviation $\mathbf{w} - \mathbf{w}'\mathbf{1}/n$, which can be interpreted as the imbalance for the intercept. These imbalances have to be smaller than 0.01 in absolute value and their average has to be smaller than 0.005. Any outlying imbalance can be reduced (in absolute value) by increasing the corresponding priority, although this has to be done with care because other imbalances may be inflated after this change.

Appendix 3 Acronyms and Capitalisations

BPD	Bronchopulmonary dysplasia
BSI	Bloodstream Infection
CHI Number	Unique numbers allocated to registered users of public health services in Scotland.
CQC	Care Quality Commission
GA	Gestational Age
HNC Number	Unique numbers allocated to registered users of public health services in Northern Ireland.
HQIP	The Healthcare Quality Improvement Partnership (HQIP) aims to promote quality improvement in patient outcomes, and in particular, to increase the impact that clinical audit, outcome review programmes and registries have on healthcare quality in England and Wales. HQIP is led by a consortium of the Academy of Medical Royal Colleges, the Royal College of Nursing and National Voices. https://www.hqip.org.uk/
IVH	Intraventricular haemorrhage
LNU	Local neonatal units (LNUs) provide neonatal care for their own catchment population, except for the sickest babies. They provide all categories of neonatal care, but they transfer babies who require complex or longer-term intensive care to a NICU, as they are not staffed to provide longer-term intensive care. Most babies over 27 weeks gestational age will usually receive their full care, including short periods of intensive care, within their LNU. Some networks have agreed variations on this policy, due to local requirements. Some LNUs provide high dependency care and short periods of intensive care for their network population. LNUs may receive transfers from other neonatal services in the network, if these fall within their agreed work pattern.
LSOA	Lower Super Output Area: an area code for a location in England.

MESH	Messaging Exchange for Social care and Health – the service used to enable users to submit lists of NHS numbers and receive lists back with the NHS numbers removed for those patients that have opted out of third party use of their data.
NEC	Necrotising enterocolitis
NHS Number	Unique numbers allocated to registered users of the three public health services in England, Wales and the Isle of Man.
NICU	Neonatal intensive care units (NICUs) are sited alongside specialist obstetric and feto-maternal medicine services and provide the whole range of medical neonatal care for their local population, along with additional care for babies and their families referred from the neonatal network. Many NICUs are co-located with neonatal surgery services and other specialised services. Medical staff in a NICU should have no clinical responsibilities outside the neonatal and maternity services.
NNAP	National Neonatal Audit Programme
ODN	Operational delivery network: In England, managed clinical networks for the coordination of neonatal critical care.
ODS	a unique code created by the Organisation Data Service within NHS Digital, and used to identify organisations across health and social care.
PHVD	Post haemorrhagic ventricular dilation
RCPCH	The Royal College of Paediatrics and Child Health (RCPCH) was founded in 1996 and now has over 17,000 members across the world. The RCPCH plays a major role in postgraduate medical education, professional standards, research and policy. https://www.rcpch.ac.uk
ROP	Retinopathy of prematurity
SCU	Special care units (SCUs) provide special care for their own local population. Depending on arrangements within their neonatal network, they may also provide some high dependency services. In addition, SCUs provide a stabilisation facility for babies who need to be transferred to a neonatal intensive care unit (NICU) for intensive or high dependency care, and they also receive transfers from other network units for continuing special care.

[1] Mood AM and Graybill FA. *Introduction to the Theory of Statistics*, 2nd ed. New York: McGraw–Hill (1963), p.408.

[2] Spiegelhalter D.J., Handling overdispersion of performance indicators, *Qual Saf Health Care*. 2005;14:347--51.

[3] The reference: Zubizarreta, J.R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association* 110, 910–922.