National Pregnancy in Diabetes Audit

Methodology Report

Published 10th October 2019
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>3</td>
</tr>
<tr>
<td>Data collection</td>
<td>4</td>
</tr>
<tr>
<td>Data quality checks</td>
<td>5</td>
</tr>
<tr>
<td>Linkage</td>
<td>6</td>
</tr>
<tr>
<td>Variables derived using other datasets and tools</td>
<td>7</td>
</tr>
<tr>
<td>- Index of Multiple Deprivation</td>
<td>7</td>
</tr>
<tr>
<td>- Birthweight centiles</td>
<td>8</td>
</tr>
<tr>
<td>Analysis</td>
<td>8</td>
</tr>
<tr>
<td>- Participation</td>
<td>8</td>
</tr>
<tr>
<td>- Data by diabetes type</td>
<td>8</td>
</tr>
<tr>
<td>- Comparisons over time</td>
<td>8</td>
</tr>
<tr>
<td>- Records with missing data</td>
<td>9</td>
</tr>
<tr>
<td>- Statistical methods used in the report</td>
<td>9</td>
</tr>
<tr>
<td>- Quartiles and 10th and 90th centiles</td>
<td>9</td>
</tr>
<tr>
<td>- Testing for a significant difference between two proportions</td>
<td>9</td>
</tr>
<tr>
<td>- Testing for a significant difference between two median values</td>
<td>9</td>
</tr>
<tr>
<td>- Confidence interval for a proportion</td>
<td>10</td>
</tr>
<tr>
<td>- Service level data</td>
<td>10</td>
</tr>
<tr>
<td>Disclosure control</td>
<td>11</td>
</tr>
<tr>
<td>Appendix 1 Data cleaning</td>
<td>12</td>
</tr>
<tr>
<td>- Rejection reasons</td>
<td>12</td>
</tr>
<tr>
<td>- Cleaning rules</td>
<td>12</td>
</tr>
</tbody>
</table>
Introduction

The National Pregnancy in Diabetes (NPID) audit is a continuous data collection measuring the quality of care and outcomes for women with pre-gestational diabetes when they become pregnant. The audit also aims to support quality improvement initiatives in this area.

All joint antenatal diabetes services in England and Wales are eligible to collect and submit audit data. The audit also receives data from the Isle of Man.

The NPID audit measures against national standards set out in the NICE (National Institute for Health and Clinical Excellence) guideline NG3 (https://www.nice.org.uk/guidance/ng3), previously NICE Clinical Guideline CG63.

The audit seeks to address three key questions:

- Were women adequately prepared for pregnancy?
- Were appropriate steps taken during pregnancy to minimise adverse outcomes to the mother?
- Did any adverse outcomes occur?

The fifth national report on pregnancies ending in 2018 was published on 10th October 2019 together with service level data aggregated for pregnancies ending in 2016, 2017 and 2018.

The audit is part of the National Diabetes Audit (NDA) programme, and is managed by the Healthcare Quality Improvement Partnership (HQIP) and funded by NHS England as part of the National Clinical Audit and Patient Outcomes Programme (NCAPOP).

The NDA is delivered by NHS Digital in partnership with Diabetes UK and is supported by Public Health England (PHE).

This document describes how the data is collected, processed and analysed before publication.
Data collection

To minimise the burden of data collection and submission the audit data is partly collected from antenatal diabetes services and partly obtained by linking the directly collected data to:

- National Diabetes Audit (NDA) data
- Hospital Episode Statistics (HES) data
- Patient Episode Database for Wales (PEDW) data
- Maternity Services Dataset (MSDS).

A full list of data items collected by the audit is available from www.digital.nhs.uk/npid

Antenatal diabetes services provide each woman eligible for inclusion in the audit with a patient information leaflet and discuss the audit with them. If the woman agrees to be included in the audit they are then provided with a consent form.

In England the data is collected under direction by NHS England, it is assumed that all women who are eligible for the audit, from all eligible sites, are entered. However this can not be verified as this information is not collected anywhere else, therefore case ascertainment cannot be calculated accurately. In Wales and the Isle of Man the antenatal diabetes team need to obtain explicit consent from each woman for the collection and submission of her data. No record is kept of how many women did not consent, so similarly to England it is not possible to calculate case ascertainment. All data collected by the NPID audit should already be stored in clinical records; no additional data should need to be collected.

Staff in antenatal diabetes services can use the NPID audit data collection form available from www.digital.nhs.uk/npid if they wish to collate the relevant data prior to entering it electronically.

The data is electronically transferred to NHS Digital by manual data entry into a secure online system called the Clinical Audit Platform (CAP). The data is entered by registered users at antenatal diabetes services. The data entry system checks each record for obvious errors and only records that pass these checks can be saved and submitted successfully.

Data can be collected and entered throughout the year. Services are asked to submit all records of pregnancies with a recorded outcome ending in a calendar year by mid-February in the following year. Any records submitted after the deadline will not be included in that year’s report but will be added to the cumulative dataset.
Data quality checks

Following the submission deadline, each service receives details of data quality checks run on their data, flagging records to review and potentially amend before a final submission deadline in late March/early April.

The number of errors in the data has reduced greatly since internal consistency checks of each record were introduced at the point of data entry in Jan 2015, (for example that the HbA1c measurements within pregnancy are between 0 and 40 weeks prior to the pregnancy end date). However some more complex errors are only detectable when the dataset is checked as a whole (for example, that the same outcome data has been attached to more than one pregnancy).

Following the final submission deadline in April, data quality checks are re-run, and either:

- Where the error would affect a lot of the analysis, the record is rejected. Across the 2014 - 2018 dataset 204 records were rejected (leaving 17,643 pregnancies).
- Where the other data in the record can still be used for analysis, the data is ‘cleaned’ by setting the value(s) assumed to be incorrect to missing.

A full list of the rejection reasons and cleaning actions is included in Appendix 1.
**Linkage**

**National Diabetes Audit data**

For each pregnancy record in the NPID data ethnicity and Lower Super Output Area (LSOA) of residence were obtained from the most recent NDA data record in the year of or prior to pregnancy. For example, for 2017 pregnancies, a match was sought in the 2017-2018 NDA, then the 2016-2017 NDA, then 2015-2016 NDA, then the 2014-2015 NDA and finally in the 2013-2014 NDA.

As mother’s diabetes type was added to the NPID data collection as a mandatory data item from 1 January 2015, this was also obtained from NDA data for the 2014 NPID audit records, again by linking to the most recent relevant NDA record. Where the diabetes type entered on the NPID system was not known, NDA linkage was also used in order to establish a known diabetes type for as many women as possible.

Year of diabetes diagnosis was obtained from the earliest NDA record available for each woman in NPID, excluding diagnosis dates that were after the start of the pregnancy or before the woman’s date of birth as recorded in the NPID data.

**Hospital Episode Statistics (HES) and Patient Episode Database for Wales (PEDW)**

Onset of labour and mode of delivery were obtained by linking the NPID audit data to HES and PEDW data.

Because HES data is released by financial year and the NPID audit reports by calendar year, only NPID pregnancies ending in 2016 or 2017 can currently be linked to available HES data.

A hospital episode record was accepted as a match when the pregnancy end date in NPID was up to 7 days earlier than the episode start date or up to 7 days after the episode end date.

The PEDW data was sparsely populated for these variables.

Details of hypoglycaemia and DKA (diabetic ketoacidosis) episodes during pregnancy were also obtained by linking to HES and PEDW data. A hospital episode record was accepted as a match when the episode started no more than 40 weeks before the estimated delivery date and on or before the pregnancy end date.

Where hypoglycaemia and DKA diagnosis codes were recorded on the same episode, this episode was counted as both an episode with hypoglycaemia and an episode with DKA. This is consistent with the method for counting complications used by the NDA.

The diagnosis codes used to determine if there was a Hypoglycaemia or Diabetic Ketoacidosis (DKA) episode during the women’s pregnancy are as follows:
Hypoglycaemia

- E16.0 Drug-induced hypoglycaemia without coma in any Diagnosis field and Y42.3 Insulin and oral hypoglycaemia (antidiabetic drugs) in a secondary diagnosis position

OR (in any Diagnosis field)

- E161 Other hypoglycaemia
- E162 hypoglycaemia, unspecified

Diabetic ketoacidosis (DKA)

The following codes in any diagnosis field

- E10.1 Type 1 diabetes mellitus with ketoacidosis
- E11.1 Type 2 diabetes mellitus with ketoacidosis
- E13.1 Other specified diabetes mellitus with ketoacidosis
- E14.1 Unspecified diabetes mellitus with ketoacidosis

Maternity Services Dataset (MSDS)

For the first time, we used MSDS to obtain parity information for NPID pregnancies. Parity is useful when using the birthweight centile tables to analyse babies born large for gestational age. MSDS tables have data fields for previous Caesarean sections, previous live births, previous stillbirths and previous losses at less than 24 weeks. These were used to create gravity and parity fields for NPID.

There was a match for over half of the NPID pregnancies between 2014 and 2018, with a better match rate in the later NPID years, when the MSDS data set was better established.

As MSDS only covers England, we tried to get similar data fields for Wales. However we did not have a key to separate pregnancies for the same woman, so linkage at pregnancy level was not possible.

Variables derived using other datasets and tools

Index of Multiple Deprivation

Local measures of deprivation are produced by the Office for National Statistics (ONS) for England and Wales separately, with the measures for each country considering different factors and using different calculation methods.

For the national NPID audit report England and Wales data have been analysed together. In order to assign a deprivation quintile to each pregnancy record based on the LSOA of residence of the mother (where this could be obtained from linking to NDA data), a ‘combined’ deprivation score was derived and ranked into quintiles. This combined score was based on an equally weighted combination of the individual scores for the employment and income indices using methodology originating from the ONS.
**Birthweight centiles**

Birthweight centiles are used to adjust the babies’ actual birthweight in line with maternal factors such as ethnicity, height and weight as well as gestational age at delivery.

Birthweight centiles were calculated for all singleton babies in the NPID dataset where the gestation at delivery and birthweight was known, using the GROW centile tool\(^1\).

A baby is described as large for gestational age (LGA) if its birthweight is above the 90\(^{th}\) centile based on gestation and maternal characteristics. In the general population, 10 per cent of babies would be expected to be above the 90\(^{th}\) centile.

**Analysis**

**Participation**

An antenatal diabetes service is counted as participating in the audit if at least one completed pregnancy record has been submitted by that service for the audit period. A full list of participating services is included in the service level excel file.

**Data by diabetes type**

The national report mainly shows data for women with Type 1 diabetes and women with Type 2 diabetes. Analysis putting these two groups together as ‘All diabetes’ can be misleading because the care pathways and physiology are often different for women with Type 1 diabetes to those with Type 2 diabetes, and the ‘All diabetes’ figure will be a reflection of the mix of diabetes types seen by the service rather than the overall quality of care.

Nevertheless, the service level reports do group together data for women with different diabetes types in order to provide some basic feedback to services that do not see a large volume of patients.

The number of women with Maturity Onset Diabetes of the Young (MODY) or ‘Other diabetes’ is too small to provide analysis for this group.

**Comparisons over time**

The analysis of the three years’ of NPID data relating to pregnancies ending in 2014, 2015 and 2016, was done for the first time in the previous report (published in 2017). At that time changes over time were not reported as there were no changes to be seen.

The analysis published in 2019 covers five years’ of NPID data and shows measures over that time period. Again, for the most part, there were few changes over the five year period.

---

\(^1\) GROW centile tool: Gardosi J, Francis A. Customised Weight Centile Calculator. GROW v8.0.4 (UK), 2019, Gestation Network, [www.gestation.net](http://www.gestation.net)
Records with missing data

Each table or chart uses all records for which the relevant data is valid and not missing in order to include as many records as possible.

This means that the denominator for percentages varies between the charts and tables.

For example, pregnancies where gestation is unknown because estimated delivery date has been removed during cleaning will be excluded from analysis where gestation is relevant, such as birthweight centiles, but will be included in other analysis where the gestation at delivery is not needed, such as whether the mother was taking 5mg folic acid prior to pregnancy.

For non-mandatory data items such as HbA1c measurements, the denominator will be much lower reflecting the number of records in which this data has not been entered.

Statistical methods used in the report

Quartiles and 10th and 90th centiles

The national report shows the variation between services for some measures, such as the percentage of women taking 5mg folic acid prior to pregnancy, and summarises the extent of variation using the median and quartile values.

Arranging all the values in order, the median is the middle value. The lower quartile is the value below which the bottom 25 per cent of data values lie and the upper quartile is the value above which the top 25 per cent of data values lie.

The 10th centile is the value below which the bottom 10 per cent of data values lie, and the 90th centile is the value above which the top 10 per cent of data values lie.

Testing for a significant difference between two proportions

Where the report explicitly compares a proportion for two groups, such as the percentage of women with Type 1 diabetes taking 5mg folic acid compared to the percentage of women with Type 2 diabetes taking 5mg folic acid, the difference between the two groups has been identified as significant by statistical testing with a p-value below 0.05 using a two-sample z-test.

Testing for a significant difference between two median values

Where the report explicitly compares a median value for two groups, such as the median BMI for women with first trimester HbA1c <48 mmol/mol and first trimester HbA1c >=48 mmol/mol, the difference between the two groups has been identified as significant by statistical testing of the distribution of the two data sets using the Mann Whitney test². The significance level used was a p-value below 0.05 on a two-tailed test.

Confidence interval for a proportion

The 95 per cent confidence intervals for the stillbirth, neonatal death and congenital anomaly rates were calculated using Byar's method as described in 'Analytical Tools for Public Health: Commonly used public health statistics and their confidence intervals' as

\[
\text{Lower limit} = \frac{0}{n} \times \left(1 - \frac{1}{9O} - \frac{1.96}{3\sqrt{O}}\right)^3
\]

\[
\text{Upper limit} = \frac{(O + 1)}{n} \times \left(1 - \frac{1}{9(O + 1)} + \frac{1.96}{3\sqrt{(O + 1)}}\right)^3
\]

where \(O\) is the observed number of events and \(n\) is the rate denominator.

Logistic regression and funnel plots

For the first time in NPID, the five years' data covering pregnancies ending in 2014-2018 was sufficiently robust to use for logistic regression modelling. Several outcomes were modelled against a range of input variables. This allowed us to see which input variables were linked with the outputs. Five years' worth of data was used to build the models and then the models were run against two years' worth of data to predict expected outcomes. These expected outcomes were compared with the actual outcomes over two years to give a standardised ratio. The expected values and ratios were presented as a funnel plot, with control limits appropriate to the volume of data. In the national report, the funnel plots were presented in order to show the range of standardised ratios across service providers. In the service level report, each organisation can see where it appears on the funnel compared to others.

Forest plots

The outputs of the SAS logistic regression models include odds ratios. These indicate the level of association between input and output variables. For the national report, the odds ratios were presented as forest plots, which allow the associations to be visualised more easily.

Service level data

The service level data published with this report includes all data for pregnancies ending in 2016-2018. The number of records for each service will in part depend on when they started participating in the audit. Some services that see a smaller number of patients per year may have more records included than a larger service because they have participated for all 3 audit years while the larger service has only joined during 2018.

The average number of pregnancy records per service is around 70.

Service level data is only published where a service has submitted 10 or more completed pregnancy records across the 3 years. As the total number of records for each service is relatively small (an average of around 70 records per service), an apparently large difference

in percentages between services may only reflect a small difference in the numerator. The service level report contains a list of participating services.

The England, Wales and Isle of Man figures in the service level report are similarly based on 3 years of data so will differ from the 2018 figures in the national report.

**Disclosure control**

Disclosure control has been applied to mitigate the risk of patient identification.

The method of disclosure control used for NPID 2018 is different from the method used for the 2016 report. We have used the standard method of suppression used across the audits which fall under the umbrella of the National Diabetes Audit.

Zeros are reported, and all numbers are rounded to the nearest 5, unless the number is 1 to 7, in which case it is rounded to ‘5’.

Rounded numbers are used to calculate percentages. At national level, this makes virtually no difference to the resultant percentages. For some services, where the numbers are small, this rounding can have a relatively large impact on the resultant percentages, but where numbers are small, percentages are volatile and should already be treated with caution.
Appendix 1 Data cleaning

Rejection reasons

A small number of records with data entry errors were rejected because the error meant that the correct complete record could not be identified or mandatory data needed to allocate a record to the correct service was missing.

The total number of records rejected from the 2014-2018 dataset was 204.

A record with pregnancy outcome data completed was rejected if:

- Two outcomes with the same pregnancy end date were attached to different pregnancy records for the same woman.
- The same baby NHS number was entered in outcome data for more than one pregnancy.
- The same baby NHS number was entered more than once in outcome data for the same pregnancy.
- Two outcomes with substantially different pregnancy end dates were attached to the same pregnancy.
- Booking or delivery hospital was missing – these are mandatory data items needed to allocate each record to a service.

Following requests from submitters, the data entry system was altered from January 2015 to allow the ‘Alive at 28 days’ question to be completed separately from the other outcome data. However, ‘Alive at 28 days’ was not completed for a number of otherwise complete and valid records which were also rejected.

Cleaning rules

Dates

The table below lists the cleaning rules applied to the NPID audit data prior to analysis in order to use as much data as possible from each record where errors remained after the data quality review by services.

Where the dates within a record were inconsistent, the pregnancy end date and pregnancy outcome have been assumed to be correct, and dates that were inconsistent with this have been set to missing.

The order of date checking means that where an estimated delivery date has been set to missing, checks which compare other dates to estimated delivery date will result in the removal of those dates.
### Table: Data cleaning rules for dates used for 2018 NPID audit report

<table>
<thead>
<tr>
<th>Data set to missing</th>
<th>Reason(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother’s date of birth</td>
<td>Age over 80 or DOB after 01/01/2003</td>
</tr>
</tbody>
</table>
| Estimated delivery date (EDD) | 1) More than 40 weeks later than pregnancy end date (negative gestation)  
2) Gestation by comparing with pregnancy end date more than 43 weeks  
3) Gestation by comparing with pregnancy end date less than 4 weeks  
4) Stillbirth at < 24 weeks  
5) Miscarriage at >=24 weeks  
6) Live birth at <20 weeks  
7) Live birth at <24 weeks with weight >1kg |
| Date of first contact with antenatal diabetes team | 1) More than 40 weeks before estimated delivery date  
2) After pregnancy end date  
3) More than 3 weeks after estimated delivery date |
| First HbA1c measurement in pregnancy and date of this measurement | 1) More than 40 weeks before estimated delivery date  
2) After pregnancy end date  
3) More than 40 weeks before pregnancy end date |
| Last HbA1c measurement in pregnancy and date of this measurement | 1) More than 40 weeks before estimated delivery date  
2) After pregnancy end date  
3) More than 40 weeks before pregnancy end date |

---

*One live birth at less than 20 weeks was retained following a request from the local service provider*
# Diabetes type

It was intended that mother’s diabetes type would be obtained by linking the NPID audit data to NDA data. However, fluctuating GP participation in the NDA and the application of patient opt-outs to the NDA data meant that this linkage was unlikely to be complete. As maternal diabetes type is an important data item for the NPID audit analysis, it was added to the NPID online data entry system for pregnancies ending on or after 1 January 2015. Whilst some services have retrospectively populated this data for some 2014 pregnancies, linkage to the NDA is still required to populate this data for as many records as possible from the earlier years.

As the NPID audit data accumulates over time, it includes second and subsequent pregnancies for many women. We would expect the woman’s diabetes type to remain the same between pregnancies. However, data quality checks revealed some possible data entry errors with different diabetes types being recorded for the same woman.

There may also be inconsistencies in diabetes type recording within the NDA data however the approach to linkage means that only one value is chosen without considering conflicts with values from other NDA years.

To maintain consistency in the dataset, the following cleaning rules have been applied to diabetes type data:

<table>
<thead>
<tr>
<th>Entered in NPID</th>
<th>Obtained from NDA</th>
<th>Used for analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>One value that is not '99 – Not specified' entered for either single pregnancy or same value for multiple pregnancies</td>
<td>Not needed</td>
<td>NPID value</td>
</tr>
<tr>
<td>One value that is not '99 – Not specified' entered (same value for one or more pregnancies) and one or more pregnancies with '99 – Not specified' or missing</td>
<td>Not needed</td>
<td>Value that is not 99 or missing applied to all pregnancies for this woman</td>
</tr>
<tr>
<td>Two different values that are not '99 – Not specified' entered for multiple pregnancies</td>
<td>Diabetes type that is not 99 or missing</td>
<td>The NPID value which matches the NDA value (ie NDA used to validate NPID)</td>
</tr>
<tr>
<td>Two different values that are not '99 – Not specified' entered for multiple pregnancies</td>
<td>Another different diabetes Type, or 99 or missing</td>
<td>NULL out the NPID values</td>
</tr>
<tr>
<td>'99 – Not specified' or missing for one or more pregnancies and no other pregnancies with a value that is not 99 or missing</td>
<td>Diabetes type that is not 99 or missing</td>
<td>NDA value</td>
</tr>
<tr>
<td>'99 – Not specified' for one or more pregnancies and no other pregnancies with a value that is not 99 or missing</td>
<td>No matches to NDA for any pregnancy</td>
<td>NPID value</td>
</tr>
</tbody>
</table>