



Validation of NPDA data

This document summaries the processes involved in the validation of NPDA data, from pre-submission through to the final quality assurance.

Stage	Process
Validation prior to submission	<ul style="list-style-type: none"> x Data collection methods and dataset changes are piloted and refined. x Data quality warnings and data completeness report generated upon each upload enable the submitting unit to identify gaps and formatting errors in their data. x Clinical lead signs form to confirm that they have reviewed their data, that no data has been submitted on behalf of patients wishing to withdraw, and that they confirm that the data are accurate. x E-mail merge to clinical leads asking to confirm name of clinical lead, unit, Trust, region. x Download most up-to-date reference files used in the analysis.
Validation of data downloaded from system	<ul style="list-style-type: none"> x The list of units within the dataset is compared against the list of units with clinical lead sign-off form. x 10 random unit level csv files are downloaded, and data are compared with the data in the extract file to make sure that the data submitted by the unit has been correctly extracted from the data capture system. x The numbdiabetes type, postcode and HbA1c are contacted when more than 50% of observations in any key variable is unknown, invalid, or missing.
Building of master file for cleaning	<ul style="list-style-type: none"> x Duplicate rows are identified (i.e. rows where every observation in the corresponding row are an exact match) and only unique rows are kept.

-
- x Blank and erroneous rows are deleted (i.e. rows where all observations are blank or contain information that is not NPDA data).
 - x Record id assigned to each visit.
 - x Questionnaire data merged with extract file.
 - Check to see which units submitted using a mixed questionnaire/csv upload method.
 - Check that units employing a mixed method approach only have questionnaire data entered subsequent to csv data entered, as new csv files should overwrite data entered via questionnaire.
 - If any questionnaire data is found , escalate to project manager and identify whether it should be incorporated or deleted.

lcal358rtif oerur(t)3 (t)2, lcal34</Mf otpxaer6.6 (3.7 (ec.3 (c 0 .3.7 (ec.3 r4 (er)12.64 500. Tm [(d)-0ed)1 (,)D.3 a(a)-5n()T o)a)- Tw y(ec.

Final quality assurance

- x A separate aggregation of regional and unit level data is undertaken and compare against the final aggregated data prepared by the analyst.