Guidance on using the National Child Measurement Programme record level dataset

England, 2015/16 school year

Published November 2016

Author: Lifestyles Statistics Team, NHS Digital
Responsible Statistician: Paul Niblett

www.digital.nhs.uk
enquiries@nhsdigital.nhs.uk

Contents

Summary Q&A

Q: What information is in this file?

A: The heights, weights and BMI for each child measured as part of the NCMP who attended a mainstream state school plus some limited demographic and geographic information.

Q: Can I use this file to identify the heaviest child in England?

A: No. The extreme weights have been suppressed to mitigate against the possibility of a child being identified (as have the extreme heights and BMI measurements).

Q: So can I use the file to say how many children are above an extreme weight such as 15 stone?

A: No as not all the weights above a certain value have been suppressed. This is because the suppression is carried out by comparing to the British 1990 growth reference dataset and suppressing measurements above the 99.995th percentile rather than using an arbitrary high weight. The growth reference dataset takes into account the child's age in days and their gender so in some cases a weight of 15 stone will be suppressed and in others it will not. The extreme low measurements below the 0.005th percentile have also been suppressed.

Q: Can you guarantee all the measurements in this file are correct?

A: The NCMP IT system contains some validation of height and weight (and BMI indirectly as it is calculated from height and weight). Specifically, values below and above certain thresholds are rejected. Other values that are within the thresholds yet are still fairly extreme are flagged for local authorities to either alter or confirm as being correct. Despite this we cannot guarantee the accuracy of all record-level submissions in a dataset of this size (over 1 million records) and the extreme values are more likely to be errors than other more typical values.

Q: So what can I do with this file?

A: The file is intended for analysts who want to carry out further research using the NCMP data. We would advise analysts to produce findings based on things like the proportion of children who are obese (or another BMI category) rather than basing findings on individual records. We would not advise carrying out analyses of weight only as weight on its own is not necessarily an indication of obesity.

Q: When should I use population thresholds and clinical thresholds?

A: Population thresholds are used for most published obesity and overweight prevalence figures, e.g. National Child Measurement Programme (NCMP) and Health Survey for England (HSE). Clinical cut-offs are recommended by NICE for use in

clinical settings with individual children and also used for NCMP parental feedback and the NHS choices BMI calculator.

Q: I need some of the other data items you collect which are not part of this file?

A: Please send an email to enquiries@hscic.gov.uk with “NCMP Data” in the subject heading outlining your requirements and we will be in touch.

Introduction

This guidance document is provided to help analysts to use the National Child Measurement Programme (NCMP) record level dataset which is made available for users to carry out their own analyses. It should be read in conjunction with the data quality statement which is available via this webpage: http://www.hscic.gov.uk/pubs/ncmpeng1516.

Certain fields have been removed and others overwritten with blanks or altered in order to ensure that children cannot be identified from the data. Further information on how this has been carried out is given within this guidance document.

Up to and including 2012/13, the record level datasets were made available to users in Microsoft Access format via the UK Data Archive: https://discover.ukdataservice.ac.uk/?q=ncmp&sf=Data+catalogue&searchType=data .

For the 2013/14 dataset onwards, it has been made available in a CSV file via the NHS Digital publication page. This change was made with a view to being more compliant with the transparency agenda by not requiring the user to apply for access to the data via the UK data archive, and also making the data available in a non-proprietary format (i.e. not requiring particular software that needs to be purchased such as Microsoft Access).

Records excluded from file

The file only contains records which were used to produce the BMI prevalence rates in the main report. Therefore records which meet the following criteria have been excluded from the file:

1. Records where the pupil has not been measured
2. Records where the measurement did not take place at a mainstream state school.

Suppression of data items

The contents of this data file have been prepared to comply with the NHS anonymisation standard[1].

Therefore, data items in the file for some records have been blanked or altered to mitigate against an individual being identified. The specific types of suppression applied follow and the number of records affected for each type is given in table 1:

1. **Suppression of extreme values** – the NCMP IT system contains some validation of height and weight (and BMI indirectly as it is calculated from height and weight). Specifically, values below and above certain thresholds are rejected. Other values that are within the thresholds yet are still fairly extreme are flagged for local authorities to either alter or confirm as being correct. Despite this we cannot guarantee the accuracy of all record-level

[1] For more details see http://www.isb.nhs.uk/documents/isb-1523/amd-20-2010/index_html.

submissions in a dataset of this size and therefore the very extreme values (height, weight or BMI above the 99.995th percentile or below the 0.005th percentile derived from British 1990 growth reference[2] have been suppressed and replaced with a value between -999 and -986. See table 2 for details of what each value means.

2. **Aggregated table suppression** – the Microsoft Excel tables which accompany the national report have undergone both primary and secondary suppression. Primary suppression consists of not showing local authority level results based on five or fewer individuals. Secondary suppression consists of removing another value so the cell which has undergone primary suppression cannot be deduced by differencing from other published totals. Records which would fall within cells which have been suppressed in Online Table 3a have had their Local Authority and Region codes, which are based on the school postcode, replaced with 'X99999999' in the record level file.
3. **Suppression of index of multiple deprivation (IMD) decile** - The IMD decile field has been blanked if the number of records with a unique combination of school year, gender, lower tier local authority and index of multiple deprivation decile is less than five. Secondary suppression has also been applied, to further decrease the risk of identifying an individual, by blanking another index of multiple deprivation decile within that combination of school year, gender and lower tier local authority which has the next fewest number of records.

Table 1: Suppression Codes

Value	Data quality issue	Number of records affected[1]
1	Suppression of low extreme values.	631
2	Suppression of high extreme values	3,192
3	Aggregated table suppression	89,260
4	Suppression of IMD decile	1,977
0	No suppression.	1,074,931

1. Records could be suppressed for more than one reason so the total sums to more than the number of records in the file. However, the table suppression was carried out before the suppression of the IMD decile which has resulted in no records being suppressed for both aggregated table suppression and IMD decile suppression.

[2] Growth monitoring with the British 1990 growth reference'. *Cole Arch Dis Child*.1997; 76: 47-49.

Table 2: Extreme Value Replacement Codes

Value	Height	Weight	BMI	Number of records affected
-999	> 99.995th percentile	> 99.995th percentile	> 99.995th percentile	23
-998	> 99.995th percentile	> 99.995th percentile		60
-997	> 99.995th percentile		> 99.995th percentile	0
-996		> 99.995th percentile	> 99.995th percentile	1,236
-995	> 99.995th percentile			506
-994		> 99.995th percentile		617
-993			> 99.995th percentile	750
-992	< 0.005th percentile	< 0.005th percentile	< 0.005th percentile	1
-991	< 0.005th percentile	< 0.005th percentile		52
-990	< 0.005th percentile		< 0.005th percentile	0
-989		< 0.005th percentile	< 0.005th percentile	39
-988	< 0.005th percentile			191
-987		< 0.005th percentile		100
-986			< 0.005th percentile	248

1. These codes have been assigned in a hierarchical order starting at the top of the table. Therefore a record which had an extremely high value for height but an extremely low value for weight and/or BMI would be assigned a replacement code for an extremely high value for height first i.e. assigned a code of -995 in this example.

Content of file

As mentioned in the introduction, not all the fields collected as part of the NCMP are being made available and some fields have been blanked or altered for some records mitigate against a child being identified. The section on suppression explains how this has been carried out.

The data items available in the file are shown in table 3.

Table 3: Contents of file

Field Name	Field Description
Ncmppseudosystemid	Unique ID code for each pupil – note that this cannot be used to link pupils over time, e.g. link their year R measurement to their year 6 measurement, in the future as it is only unique within the collection year and not over time.
Genderdescription	Sex of pupil.
Ageinmonths	Age of pupil (in months).
Schoolyear	School Year of pupil - derived from child age (R: Reception, 6: Year 6).
height[2]	Height of pupil (in cm).
heightzscore[2]	Height z score - derived from British 1990 growth reference, using Age, Sex and Height fields[1].
heightpscore[2]	Height centile - derived from British 1990 growth reference, using Age, Sex and Height fields[1].
weight[2]	Weight of pupil (in kg).
weightzscore[2]	Weight z score - derived from British 1990 growth reference, using Age, Sex and Weight fields[1].
weightpscore[2]	Weight centile - derived from British 1990 growth reference, using Age, Sex and Weight fields[1].
bmi[2]	BMI of pupil in kg/m^2 - derived from height and weight.
bmizscore[2]	BMI z score - derived from British 1990 growth reference, using Age, Sex and BMI fields[1].
bmipscore[2]	BMI centile - derived from British 1990 growth reference, using Age, Sex and BMI fields[1].
bmipopulationcategory	BMI classification to UK90 population monitoring centiles (above 85th/95th centiles for overweight and obese and below 2nd centile for underweight[1]).
bmiclinicalcategory	BMI classification to UK90 clinical centiles (BMI z-score over 4/3 for overweight and over 2 for obese and below -2 for underweight[1]).
schooltier1localauthoritycode[3]	ONS code for Local Authority (Tier1) – derived from postcode of school.
schooltier2localauthoritycode[3]	ONS code for Local Authority (Tier2) – derived from postcode of school.
schoolgovernementofficeregion[3]	ONS code for Government Office Region - derived from postcode of school.
Schoolindexofmultipledepriv	The decile that the Index of Multiple Deprivation (IMD) 2015 score falls into - derived from LSOA of school (1=1st decile (most deprived) to 10=10th decile (least deprived)).
pupilschooldistancebanded	The straight line distance between the pupil postcode and school postcode (in km) banded.
suppress_record_high	Record has extreme high values and has had height, weight, bmi and associated p-scores and z-scores removed.
suppress_record_low	Record has extreme low values and has had height, weight, bmi and associated p-scores and z-scores removed.
suppress_table	Record has had LA and region codes removed as it falls within a small cell in the published online table 3a.
suppress_imd	Record has had the IMD decile removed because the combination of school year, gender, lower tier local authority and IMD decile is less than five.

1. Growth monitoring with the British 1990 growth reference'. Cole *Arch Dis Child.*1997; 76: 47-49.
2. Some extreme values have been masked
3. Table 3A in the NCMP national report can be used to create a lookup table for schooltier1localauthority, schooltier2localauthority and schoolgovernmentoffice region.

Recommendations for NCMP analysis

Users of the data may also be interested in guidance documents written by Public Health England (PHE) on using record level NCMP data. These are available at: http://www.noo.org.uk/NCMP/analytical_guidance.

Users should not base analyses or articles on individual records with extreme values and should instead concentrate on findings from the whole dataset such as the proportion of children in different BMI categories.

This is because we cannot guarantee the accuracy of all record-level submissions in a dataset of this size, although we do apply validation checks throughout the NCMP data collection process and work closely with local authorities to help ensure the records they submit are accurate.

Users are urged not to use weight in isolation and instead to use the BMI p-score as the heaviest children may not necessarily be the most obese children.

Wherever possible, local analysis should be checked against the figures published by the NHS Digital and PHE to ensure consistency. However, be aware that you will not be able to match any published analyses using the index of multiple deprivation decile as some of these values have been suppressed as explained previously.

Confidence limits are published in the Excel data tables provided by NHS Digital and PHE and these should be used when making comparisons between areas or monitoring change over time. Users of this file should also apply confidence limits or statistical tests to their own analyses. The methodology for doing this is in annex E in the appendices accompanying the national report[3].

Any publications using NCMP data should clearly state whether the population or clinical thresholds have been used to derive obesity and overweight prevalence figures.

Conditions of Use

Any published analysis should include ‘NHS Dgitial’ as the data source for NCMP data.

What to do if this extract isn’t suitable for your needs

It is a requirement that the extract complies with the NHS anonymisation standard. This has been achieved by removing some data items and changing others in this file to mitigate against an individual being identified. User feedback was sought before deciding on the fields to include and the suppression methods to use in the extract.

For example, it was found that users thought the index of multiple deprivation decile field was more important than the ethnicity field. Consequently the ethnicity of the pupil has not been included and it was decided to blank the index of multiple deprivation decile rather than say, the lower tier local authority code.

[3] http://www.content.digital.nhs.uk/catalogue/PUB22269/nati-chil-meas-prog-eng-2015-2016-app.pdf

If you would like to provide feedback on the extract, please email enquiries@nhsdigital.nhs.uk and state “NCMP Data Extract” in the subject line. We will take user feedback into account when constructing next year’s extract and deciding on fields to include and suppression methodology.

Feedback

If you have any comments on this document or the contents of the record level dataset then please email enquiries@nhsdigital.nhs.uk quoting “NCMP Data Extract” in the subject field.

We are keen to understand how our users use our data so we would be interested to hear what you used this record level dataset for. Feedback can be sent using the same contact details.

Future user engagement

Any contact details gathered as correspondence in relation to this file may be used by NHS Digital for future user engagement about the NCMP data such as making changes to the dataset or changes to the published outputs. If you do not wish your contact details to be used in this way then please include the following text in any email correspondence:

“I do not wish my contact details to be used for any future user engagement on NCMP data”.

Information and technology
for better health and care

www.digital.nhs.uk

0300 303 5678

enquiries@nhsdigital.nhs.uk

@nhsdigital

This publication may be requested
in large print or other formats.

Published by NHS Digital, part of the Government Statistical Service

NHS Digital is the trading name of the
Health and Social Care Information Centre.