



Health & Social Care
Information Centre



Public Health
England



Guidance on using the National Child Measurement Programme record level dataset:

England, 2014/15 school year

Published January 2016



**We are the trusted
national provider of
high-quality information,
data and IT systems for
health and social care.**

www.hscic.gov.uk

enquiries@hscic.gov.uk

 **[@hscic](https://twitter.com/hscic)**

Author: Lifestyles Statistics Team, Health and
Social Care Information Centre

Responsible statistician: Paul Niblett, Section Head

Version: V1.0

Date of publication: January 2016

Contents

Introduction	4
Records excluded from file	4
Suppression of data items	4
Content of file	6
Recommendations for NCMP analysis	8
Conditions of Use	8
What to do if this extract isn't suitable for your needs	8
Feedback	9
Future user engagement	9

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 also available via this webpage:

<http://www.hscic.gov.uk/pubs/ncmpeng1415>

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 file¹ via the HSCIC publication page. This change was made with a view to being more compliant with the transparency agenda² by not requiring the user to apply for access to the data via the UK data archive³, 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 state school.

Suppression of data items

The contents of this data file have been prepared to comply with the NHS anonymisation standard⁴.

Therefore, data items in the file for a small number of 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

¹ The CSV file sits within a ZIP file in order to minimise download time for users. The zip file also contains this guidance document.

² For more information on the transparency agenda see <https://www.gov.uk/government/policies/improving-the-transparency-and-accountability-of-government-and-its-services>

³ One weakness of this approach is that we will not automatically gather information on the uses made of the data which was part of the application process. However, we are requesting users to still tell us how they use the data and details on how to do this are given in the feedback section of this document.

⁴ For more details see <http://www.isb.nhs.uk/documents/isb-1523/amd-20-2010/index.html>.

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 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 reference⁵ 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 fewer than five 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 the next smallest index of multiple deprivation decile within that combination of school year, gender and lower tier local authority.

Table 1: Suppression Codes

Value	Data quality issue	Number of records affected ¹
1	Suppression of low extreme values.	655
2	Suppression of high extreme values	2,817
3	Aggregated table suppression	76,249
4	Suppression of IMD decile	1,887
0	No suppression.	1,060,307

Notes:

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 these reasons.

⁵ 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	> 99.995 th percentile	> 99.995 th percentile	> 99.995 th percentile	21
-998	> 99.995 th percentile	> 99.995 th percentile		62
-997	> 99.995 th percentile		> 99.995 th percentile	0
-996		> 99.995 th percentile	> 99.995 th percentile	1,113
-995	> 99.995 th percentile			453
-994		> 99.995 th percentile		574
-993			> 99.995 th percentile	594
-992	< 0.005 th percentile	< 0.005 th percentile	< 0.005 th percentile	1
-991	< 0.005 th percentile	< 0.005 th percentile		56
-990	< 0.005 th percentile		< 0.005 th percentile	0
-989		< 0.005 th percentile	< 0.005 th percentile	46
-988	< 0.005 th percentile			195
-987		< 0.005 th percentile		107
-986			< 0.005 th percentile	250

Notes:

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 in order to avoid identification of individual children. The section on suppression explains how this has been carried out.

The data items 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.
bmi zscore ²	BMI z score - derived from British 1990 growth reference, using Age, Sex and BMI fields ¹ .
bmi pscore ²	BMI centile - derived from British 1990 growth reference, using Age, Sex and BMI fields ¹ .
bmi populationcategory	BMI classification to UK90 population monitoring centiles (85th/95th centiles for overweight and obese), plus 2nd centile for underweight ¹ .
bmi clinicalcategory	BMI classification to UK90 clinical centiles (approx 91st/98th centiles for overweight and obese), plus 2nd centile for underweight ¹ .
schooltier1localauthority ³	ONS code for Local Authority (Tier1) – derived from postcode of school.
schooltier2localauthority ³	ONS code for Local Authority (Tier2) – derived from postcode of school.
schoolgovernmentofficeregion ³	Government Office Region - derived from postcode of school.
Schoolindexofmultipledepriv	The decile that the Index of Multiple Deprivation (IMD) 2010 score falls into - derived from LSOA of school (0: not available, 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-values and z-scores removed.
suppress_record_low	Record has extreme low values and has had height, weight, bmi and associated p-values 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 as it falls within a cell with few records in the combination of school year, gender, lower tier local authority and IMD decile.

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. For more information about how BMI is classified in children please refer to the following document produced by the National Obesity Observatory (NOO): http://www.noo.org.uk/uploads/doc/vid_11762_classifyingBMlinchildren.pdf

Wherever possible, local analysis should be checked against the figures published by the HSCIC and PHE Obesity Knowledge & Intelligence team (K&I) 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 the HSCIC and PHE Obesity K&I and these should be used when making comparisons between areas or monitoring change over time. Users of the NCMP dataset should also apply confidence limits or statistical tests to their own analyses. The methodology for doing this is in an annex in the national report.

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 'The Health and Social Care Information Centre (HSCIC)' 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 collected 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.

If you would like to provide feedback on the extract, please email enquiries@hscic.gov.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@hscic.gov.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 the HSCIC 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”.

**Published by the Health and Social Care Information Centre
Part of the Government Statistical Service**

Responsible Statistician

Paul Niblett, Section Head Lifestyles

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

For further information

www.hscic.gov.uk

0300 303 5678

enquiries@hscic.gov.uk

Copyright © 2016 Health and Social Care Information Centre. All rights reserved.

This work remains the sole and exclusive property of the Health and Social Care Information Centre and may only be reproduced where there is explicit reference to the ownership of the Health and Social Care Information Centre.

This work may be re-used by NHS and government organisations without permission.