## Data and data dictionary

Version 1.0

Date: 14 December 2020

### Data anonymisation

Data are provided according to the study data disclosure plan (version 1.1, 6 December 2020), lodged with the Health Research Authority (HRA).

The HRA advised that sharing an anonymised dataset from the study was possible despite the lack of consent from participants if this was in the public interest and if key data governance principles were followed.

The purpose of deposition is a commitment to scientific transparency and to allow other researchers to replicate findings to the greatest extent possible compatible with sharing of anonymised data. A separate principle of data minimisation has also been respected, i.e. deposit of the smallest amount of data that achieves the objectives whilst protecting participants. Combining these two principles requires us to deposit the minimum data that allows replication of the most findings in our analyses whilst absolutely preserving the identity of participants.

The process of anonymisation and deposit was guided by the following documents.

1. Elliot et al. The Anonymization Decision-Making Framework. UKAN, University of Manchester, 2016. <https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf>
2. Microdata Handling and Security: Guide to Good Practice. UK Data Archive, University of Essex Version 6.0, 9 Feb 2017 <https://ukdataservice.ac.uk/media/604725/cd171-microdatahandling.pdf>
3. V. Van den Eynden, K. Schreckenberg, L. Corti. Sharing social data in multidisciplinary, multistakeholder research: Best practice guide for researchers. UK Data Service and ESPA, ESPA Working Paper Series No: 001 / December 2015; ISSN 20589875. <http://www.espa.ac.uk/files/espa/Sharing%20social%20data.pdf>
4. Tsai AC, Kohrt BA, Matthews LT, et al. Promises and pitfalls of data sharing in qualitative research. *Soc Sci Med*. 2016;169:191-198. doi:10.1016/j.socscimed.2016.08.004
5. UCL guidance on Anonymisation and Pseudonymisation

 <https://www.ucl.ac.uk/data-protection/guidance-staff-students-and-researchers/practical-data-protection-guidance-notices/anonymisation-and>

1. Branson J, Good N, Chen JW, Monge W, Probst C, El Emam K. Evaluating the re-identification risk of a clinical study report anonymized under EMA Policy 0070 and Health Canada Regulations. Trials. 2020;21(1):200. Published 2020 Feb 18. doi:10.1186/s13063-020-4120-y

We further conducted a Motivated Intruder Test, as recommended by Trevor Peacock (Head of Information Security at University College London) and also following Branson et al. (2020). Testing identified there were large amounts of data in the public domain from news media and social media identifying individuals in the UK who had similar treatment to that given in the study during approximately the same dates. The motivated intruder test was not successful in de-identifying participants.

All participants in the study have been notified in writing that these data are being deposited and offered the opportunity to withdraw their data. At the time of deposit (December 2020), no participants had noted they wished to withdraw their data.

The data from the study have been deposited according to the following guidelines:

*1. All personal identification data and potentially sensitive personal and demographic data have been removed.*

Items to be removed will include:

a. personal identifiers e.g. name, date of birth

b. ethnicity and family structure

c. all data relating to recruitment or times or dates of clinic attendances.

d. age: note the restricted age-range (12-15 at recruitment) meant that obscuring of age through age-banding was not possible.

e. birth-assigned sex.

Note that a new pseudo identifier for each participant will be created in the deposited dataset.

*2. All treatment and medical information have been removed*

There is a risk that data on timing of clinic or investigation visits, treatment decisions, side effects and fertility issues may lead to identification. Further, all subjects received GnRHa at (essentially) the same dose and thus treatment data are not necessary to understand the outcomes of treatment.

We therefore removed the following data:

1. Timing of clinic visits or medical investigations
2. Treatment decisions, time off treatment, any data on fertility issues. This includes the decision to start cross-sex hormones (as 1 subject did not start cross-sex hormones)
3. karyotype, hormone levels and other venepuncture data
4. data on side effects / adverse events

Note that all timing in the deposited data are standardised to baseline and 12, 36 and 48 months in follow-up, conforming to the analyses undertaken and removing risk of identification through timing.

*3. Data from the semi-structured interviews will not be deposited*

This is because of the higher risk of disclosure and identification from qualitative data (Tsai, 2016).

*4. Outcome data will be deposited as standardised data only*

Outcome data are released as standardized data i.e. data standardised for age and/or sex. Data used in the project were standardised in this way where possible. This is to avoid potential identification through more extreme raw scores, as standardised scores are one step removed from the individual. We will not release data that cannot be standardised i.e. where appropriate normative and standardisation data do not exist. We will not deposit item-level data (i.e. the items that make up psychological scales) as these hold potential for re-identification (e.g. participant personal stories and medical histories may be available online through the media or social media).

Data deposited consist of:

1. standardised psychological scale scores i.e. t-scores and z-scores
2. standardised bone mineral density z-scores
3. standardised height, weight and BMI z-scores

*5. Variables with n<8 have been removed*

In our published paper, we did not report data from small cells i.e. where sample size was less than 8. For the same reasons we do not deposit these variables. Clearly they are not required for replication of analyses.

### File

The dataset is contained in a single Stata file, version 11/12

### Version

*GD public dataset deposited v1.0.dta* - deposited on 14 December 2020.

### Variables

All variables are numeric. The label for each variable is self-explanatory. In terms of time, data are provided for baseline (before treatment), and follow-up at 12, 24 and 36 months. These are shown as follows:

* Baseline: before starting treatment. All baseline variables have the suffice ‘b’
* Follow-up: This occurred at 12, 24 and 36 months. Each of these variables has the suffix 12m, 24m or 36m as appropriate.

For details of the included data and scales see our deposited preprint (<https://www.medrxiv.org/content/10.1101/2020.12.01.20241653v1>) or published paper (to be updated when published).

Variables deposited:

|  |  |  |
| --- | --- | --- |
| *variable label* |  | *variable* |
| pseudo identifier |  | id |
| spine BMD z-score baseline |  | spinezb |
| hip BMD z-score baseline |  | hipzb |
| spine BMD z-score 12 months |  | spinez12m |
| hip BMD z-score 12 months |  | hipz12m |
| spine BMD z-score 24 months |  | spinez24m |
| hip BMD z-score 24 months |  | hipz24m |
| spine BMD z-score 36 months |  | spinez36m |
| hip BMD z-score 36 months |  | hipz36m |
| spine height-adjusted BMD z-score baseline |  | spinehazzb |
| spine height-adjusted BMD z-score 12 months |  | spinehazz12m |
| spine height-adjusted BMD z-score 24 months |  | spinehazz24m |
| spine height-adjusted BMD z-score 36 months |  | spinehazz36m |
| hip height-adjusted BMD z-score baseline |  | hiphazzb |
| hip height-adjusted BMD z-score 12 months |  | hiphazz12m |
| hip height-adjusted BMD z-score 24 months |  | hiphazz24m |
| hip height-adjusted BMD z-score 36 months |  | hiphazz36m |
| height z-score baseline |  | zhtb |
| weight z-score baseline |  | zwtb |
| BMI z-score baseline |  | zbmib |
| height z-score 12 months |  | zht12m |
| weight z-score 12 months |  | zwt12m |
| BMI z-score 12 months |  | zbmi12m |
| height z-score 24 months |  | zht24m |
| weight z-score 24 months |  | zwt24m |
| BMI z-score 24 months |  | zbmi24m |
| height z-score 36 months |  | zht36m |
| weight z-score 36 months |  | zwt36m |
| BMI z-score 36 months |  | zbmi36m |
| YSR Total t-score baseline |  | YSR\_b\_Total\_Tscore |
| YSR External t-score baseline |  | YSR\_b\_External\_Tscore |
| YSR Internal t-score baseline |  | YSR\_b\_Internal\_Tscore |
| CBCL Total t-score baseline |  | CBCL\_b\_Total\_Tscore |
| CBCL External t-score baseline |  | CBCL\_b\_External\_Tscore |
| CBCL Internal t-score baseline |  | CBCL\_b\_Internal\_Tscore |
| YSR Total t-score 12 months |  | YSR\_12m\_Total\_Tscore |
| YSR External t-score 12 months |  | YSR\_12m\_External\_Tscore |
| YSR Internal t-score 12 months |  | YSR\_12m\_Internal\_Tscore |
| CBCL Total t-score 12 months |  | CBCL\_12m\_Total\_Tscore |
| CBCL External t-score 12 months |  | CBCL\_12m\_External\_Tscore |
| CBCL Internal t-score 12 months |  | CBCL\_12m\_Internal\_Tscore |
| CBCL Internal t-score 24 months |  | CBCL\_Internal\_Tscore\_24m |
| CBCL External t-score 24 months |  | CBCL\_External\_Tscore\_24m |
| CBCL Total t-score 24 months |  | CBCL\_Total\_Tscore\_24m |
| CBCL Internal t-score 36 months |  | CBCL\_Internal\_Tscore\_36m |
| CBCL External t-score 36 months |  | CBCL\_External\_Tscore\_36m |
| CBCL Total t-score 36 months |  | CBCL\_Total\_Tscore\_36m |
| YSR Internal t-score 24 months |  | YSR\_Internal\_Tscore\_24m |
| YSR External t-score 24 months |  | YSR\_External\_Tscore\_24m |
| YSR Total t-score 24 months |  | YSR\_Total\_Tscore\_24m |
| YSR total z-score Australia reference baseline |  | zaustYSRb |
| YSR total z-score Netherlands reference baseline |  | znethYSRb |
| YSR total z-score Australia reference 12 months |  | zaustYSR12m |
| YSR total z-score Netherlands reference 12 months |  | znethYSR12m |
| YSR total z-score Australia reference 24 months |  | zaustYSR24m |
| YSR total z-score Netherlands reference 24 months |  | znethYSR24m |
| Kidscreen Parent subscale 1 b |  | KSParent\_b\_subscale1 |
| Kidscreen Parent subscale 2 b |  | KSParent\_b\_subscale2 |
| Kidscreen Parent subscale 3 b |  | KSParent\_b\_subscale3 |
| Kidscreen Parent subscale 4 b |  | KSParent\_b\_subscale4 |
| Kidscreen Parent subscale 5 b |  | KSParent\_b\_subscale5 |
| Kidscreen Parent subscale 6 b |  | KSParent\_b\_subscale6 |
| Kidscreen Parent subscale 7 b |  | KSParent\_b\_subscale7 |
| Kidscreen Parent subscale 8 b |  | KSParent\_b\_subscale8 |
| Kidscreen Parent subscale 9 b |  | KSParent\_b\_subscale9 |
| Kidscreen Parent subscale 10 b |  | KSParent\_b\_subscale10 |
| Kidscreen Parent subscale 1 12m |  | KSParent\_12m\_subscale1 |
| Kidscreen Parent subscale 2 12m |  | KSParent\_12m\_subscale2 |
| Kidscreen Parent subscale 3 12m |  | KSParent\_12m\_subscale3 |
| Kidscreen Parent subscale 4 12m |  | KSParent\_12m\_subscale4 |
| Kidscreen Parent subscale 5 12m |  | KSParent\_12m\_subscale5 |
| Kidscreen Parent subscale 6 12m |  | KSParent\_12m\_subscale6 |
| Kidscreen Parent subscale 7 12m |  | KSParent\_12m\_subscale7 |
| Kidscreen Parent subscale 8 12m |  | KSParent\_12m\_subscale8 |
| Kidscreen Parent subscale 9 12m |  | KSParent\_12m\_subscale9 |
| Kidscreen Parent subscale 10 12m |  | KSParent\_12m\_subscale10 |
| Kidscreen Child subscale 1 b |  | KSChild\_b\_subscale1 |
| Kidscreen Child subscale 2 b |  | KSChild\_b\_subscale2 |
| Kidscreen Child subscale 3 b |  | KSChild\_b\_subscale3 |
| Kidscreen Child subscale 4 b |  | KSChild\_b\_subscale4 |
| Kidscreen Child subscale 5 b |  | KSChild\_b\_subscale5 |
| Kidscreen Child subscale 6 b |  | KSChild\_b\_subscale6 |
| Kidscreen Child subscale 7 b |  | KSChild\_b\_subscale7 |
| Kidscreen Child subscale 8 b |  | KSChild\_b\_subscale8 |
| Kidscreen Child subscale 9 b |  | KSChild\_b\_subscale9 |
| Kidscreen Child subscale 10 b |  | KSChild\_b\_subscale10 |
| Kidscreen Child subscale 1 12m |  | KSChild\_12m\_subscale1 |
| Kidscreen Child subscale 2 12m |  | KSChild\_12m\_subscale2 |
| Kidscreen Child subscale 3 12m |  | KSChild\_12m\_subscale3 |
| Kidscreen Child subscale 4 12m |  | KSChild\_12m\_subscale4 |
| Kidscreen Child subscale 5 12m |  | KSChild\_12m\_subscale5 |
| Kidscreen Child subscale 6 12m |  | KSChild\_12m\_subscale6 |
| Kidscreen Child subscale 7 12m |  | KSChild\_12m\_subscale7 |
| Kidscreen Child subscale 8 12m |  | KSChild\_12m\_subscale8 |
| Kidscreen Child subscale 9 12m |  | KSChild\_12m\_subscale9 |
| Kidscreen Child subscale 10 12m |  | KSChild\_12m\_subscale10 |
| Kidscreen Child subscale 1 24m |  | KSChild\_24m\_subscale1 |
| Kidscreen Child subscale 2 24m |  | KSChild\_24m\_subscale2 |
| Kidscreen Child subscale 3 24m |  | KSChild\_24m\_subscale3 |
| Kidscreen Child subscale 4 24m |  | KSChild\_24m\_subscale4 |
| Kidscreen Child subscale 5 24m |  | KSChild\_24m\_subscale5 |
| Kidscreen Child subscale 6 24m |  | KSChild\_24m\_subscale6 |
| Kidscreen Child subscale 7 24m |  | KSChild\_24m\_subscale7 |
| Kidscreen Child subscale 8 24m |  | KSChild\_24m\_subscale8 |
| Kidscreen Child subscale 9 24m |  | KSChild\_24m\_subscale9 |
| Kidscreen Child subscale 10 24m |  | KSChild\_24m\_subscale10 |
| Kidscreen Parent subscale 1 24m |  | KSParent\_24m\_subscale1 |
| Kidscreen Parent subscale 2 24m |  | KSParent\_24m\_subscale2 |
| Kidscreen Parent subscale 3 24m |  | KSParent\_24m\_subscale3 |
| Kidscreen Parent subscale 4 24m |  | KSParent\_24m\_subscale4 |
| Kidscreen Parent subscale 5 24m |  | KSParent\_24m\_subscale5 |
| Kidscreen Parent subscale 6 24m |  | KSParent\_24m\_subscale6 |
| Kidscreen Parent subscale 7 24m |  | KSParent\_24m\_subscale7 |
| Kidscreen Parent subscale 8 24m |  | KSParent\_24m\_subscale8 |
| Kidscreen Parent subscale 9 24m |  | KSParent\_24m\_subscale9 |
| Kidscreen Parent subscale 10 24m |  | KSParent\_24m\_subscale10 |