It is made available under a [CC-BY-NC-ND 4.0 International license](http://creativecommons.org/licenses/by-nc-nd/4.0/) . **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. medRxiv preprint doi: [https://doi.org/10.1101/2024.08.19.24312213;](https://doi.org/10.1101/2024.08.19.24312213) this version posted August 20, 2024. The copyright holder for this preprint

K.J. Low^{1.2}, J. Foreman³, R.J. Hobson⁴, H. Kwuo¹, E.Martinez-Cayuelas⁵, B. Almoguera^{6,7}, P. Marin-Reina⁸, S.G. Caraffi⁹, L. Garavelli⁹, E. Woods^{10.11}, M. Balasubramanian ^{10.11}, A. Bayat^{12.13.14}, C.W K.J. Low^{1,2}, J. Fo
Reina⁸, S.G. Car
Ockeloen¹⁵, C.N
1 Centre for Ac $\begin{array}{c} \n\text{I} \quad \text{I} \quad \$ K.J. Low^{4,2}, J. Foreman³
Reina⁸, S.G. Caraffi⁹, L. (
Ockeloen¹⁵, C.M. Wrigh
1 Centre for Academic
2 Department of Clinica
3 European Molecular I
Campus, Hinxton, Caml , R.J. Hobson⁻
Garavelli⁹, E. \
t¹⁶, H.V. Firth¹⁶
Child Health, I
al Genetics, UI
Biology Labora , H. Kwuo¹,
Woods^{10,11},
^{3,17}, T.J. Col
University
HBW NHS
atory, Euro , E.Martinez-Cayuelas
M. Balasubramanian ¹
le¹⁸
of Bristol, Bristol, UK
Trust, Bristol, UK
pean Bioinformatics Ir , B. Almoguera^{o, 7}, P. Marin-
^{0,11} , A. Bayat^{12,13,14}, C.W.
stitute, Wellcome Genome Reina⁸, S.G. Caraffi⁹, L. Garavelli⁹, E. Woods^{10,11}, M. Balasubramanian^{10,11}, A. Bayat^{12,13,14}, C.W.

Reina
Ockelc
1 Cent
2 Depa
3 Euro
Campu , S.G. Caraffi $^{\circ}$
)en¹⁵, C.M. W
re for Acader
artment of Cli
pean Molecu
us, Hinxton, C , L. Garavelli⁹
right¹⁶, H.V. F
mic Child Hea
nical Genetic
lar Biology La
ambridge, CE , E. Woods^{20,11}, M. Balasubramanian ^{10,11}
Firth^{3,17}, T.J. Cole¹⁸
Ith, University of Bristol, Bristol, UK
s, UHBW NHS Trust, Bristol, UK
aboratory, European Bioinformatics Instit
10 1SD, UK , A. Bayat
12, C.W.
14, Wellcome Genom
2 Ockeloen¹⁹, C.M. Wright¹⁹, H.V. Firth^{3,17}, T.J. Cole¹⁸
1 Centre for Academic Child Health, University of B
2 Department of Clinical Genetics, UHBW NHS Trus
3 European Molecular Biology Laboratory, Europea
Campus, H $\frac{1}{2}$ $\frac{1}{2}$ 2 Department of Clinical Genetics, UHBW NHS Trust, Bristol, UK
3 European Molecular Biology Laboratory, European Bioinformatics
Campus, Hinxton, Cambridge, CB10 1SD, UK
4 DDD team, Wellcome Sanger Institute, Hinxton, Cambr

2 Department of Cambridge Laboratory, European Bioinforma
2 Buropean Molecular Biology Laboratory, European Bioinforma
2 Campus, Hinxton, Cambridge, CB10 1SD, UK
4 DDD team, Wellcome Sanger Institute, Hinxton, Cambridge, U 3 European Molecular Biology Laboratory, European Biology Campus, Hinxton, Cambridge, CB10 1SD, UK
3 DDD team, Wellcome Sanger Institute, Hinxton, Cambridge, UK
5 Department of Pediatrics, Hospital Universitario Fundación Example 18 and Mellcome Sanger Institute, Hin:

5 Department of Pediatrics, Hospital Univers

6. Department of Genetics and Genomics, Fu

Research Institute-Fundacion Jimenez Diaz, L

Madrid, Spain.

7. Center for Biomedic 5 Department of Pediatrics, Hospital Universitario Fundación Jimé
6. Department of Genetics and Genomics, Fundacion Jimenez Dia
Research Institute-Fundacion Jimenez Diaz, Universidad Autonom
Madrid, Spain.
7. Center for Bi 5 Department of Genetics and Genomics, Fundacion Jimenez Diaz University Hospital, Hea
Research Institute-Fundacion Jimenez Diaz, Universidad Autonoma de Madrid (IIS-FJD, UAT
Madrid, Spain.
7. Center for Biomedical Network Research Institute-Fundacion Jimenez Diaz, Universidad Autonoma de Madrid (IIS-FJD, UAM),
Madrid, Spain.
7. Center for Biomedical Network Research on Rare Diseases (CIBERER), Madrid, Spain.
8. Dysmorphology and Neonatology Research Institute Constant Institute Constant Institute Constant Institute (IIC-FJC)
T. Center for Biomedical Network Research on Rare Diseases (CIBERER), Madrid, Spain.
8. Dysmorphology and Neonatology Unit. Hospital Uni

Madrid, Spain.
7. Center for Bi
8. Dysmorpholo
9. Medical Gen
10 Division of C
11 Department
12 Department

2. Center for Biomeral Americal Americal Americal Constant, probably spaintal Research on Research on Research
19. Medical Genetics Unit, Azienda USL-IRCCS di Reggio Emilia, 42122 Reggio Emilia, Italy
10 Division of Clinic

11 Department of Clinical Genetics, Sheffield Children's Hospital, She
12 Department of Regional Health Research, University of Southern I
TE: This preprint reports new research that has not been certified by peer review a

9. Medical Genetics Unit, Azienda USL-IRCCS di Reggio Emilia, 42122 Reggio Emilia, Italy
10 Division of Clinical Medicine, University of Sheffield, Sheffield, UK
11 Department of Clinical Genetics, Sheffield Children's Hos 9. Medical Genetics University of Sheffield, Sheffield, UK
10 Division of Clinical Medicine, University of Sheffield, Sheffield, UK
11 Department of Clinical Genetics, Sheffield Children's Hospital, Sheffield UK
12 Departm 11 Department of Regional Health Research, University of Southern Denmark,
12 Department of Regional Health Research, University of Southern Denmark,
TE: This preprint reports new research that has not been certified by pe

 $\mathbf{1}$

13 Department of Epilepsy Genetics and Personalized Medicine, Danish Epilepsy Center, Dianalund, Imark
Department of Drug Design and Pharmacology, University of Copenhagen, Copenhagen, Denmark
Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands
Department of Human Nutrition, Scho

14 Department of Drug Design and Pharmacology, University of Copenhagen, Copenhagen, Denmark

14 Depart
14 Depart
15 Depart
16 Depart
Glasgow,
17 Depart Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands
Department of Human Nutrition, School of Medicine, Dentistry and Nursing, University of
sgow, Glasgow, UK
Department of Clinical Ge

16 Department of Human Nutrition, School of Medicine, Dentistry and Nursing, University of
Glasgow, Glasgow, UK
17 Department of Clinical Genetics, Addenbrookes Hospital, Cambridge, UK
18 UCL Great Ormond Street Institute

16 Glasgow, Glasgow, UK
17 Department of Clinical Genetics, Addenbrookes Hospital, Cambridge, UK
18 UCL Great Ormond Street Institute of Child Health, London, UK
Corresponding author
Dr Karen Low gow, Glasgow, Black
17 Department of Clini
18 UCL Great Ormond
Corresponding author
Dr Karen Low
Karen Low

18 UCL Great Ormond Street Institute of Child Health, London, UK
18 UCL Great Ormond Street Institute of Child Health, London, UK
Corresponding author
Dr Karen Low @bristol.ac.uk
00441174553328

 $\frac{1}{2}$

18 UCL Great Ormondon
18 Uclean Ormondon
18 National Street Institute of Child Health
18 National Health
18 National Health, Northern Street Institute of Child Health
18 National Meanwhile Corresponding author
Dr Karen Low
<u>Karen low@bristol.ac.u</u>
00441174553328 Dr Karen Low

Maria 2004
00441174553328
004

Introduction

Methods

growth charts would be useful but require large samples to construct them using the conventio
LMS method.
Methods
We transformed anthropometry to British 1990 reference z-scores for 328 UK and 264 internation
probands wit LMS method.
Methods
We transformed anthropometry to British 1990 reference z-scores for 328 UK and 264 international
probands with ANKRD11, ARID1B, ASXL3, DDX3X, KMT2A or SATB2-related disorders, and modelled
mean and stan Methods
We transform
probands with
mean and stan
Back-transform
Results
The resulting :

Results

probands with *ANKRD11, ARID1B, ASXL3, DDX3X, KMT2A* or *SATB2*-related disorders, and modelled
mean and standard deviation (SD) of the z-scores as gene-specific linear age trends adjusted for sex.
Back-transforming the me probands with ANKRD11, ARD15, ASXL3, DDX3X, MMT2A or SATD2-related disorders, and inodelical
mean and standard deviation (SD) of the z-scores as gene-specific linear age trends adjusted for sex.
Back-transforming the mean mean and standard deviation (see your deviation galaxies) of the z-spectrum on a
Back-transforming the mean ±2 SD lines provided gene-specific median, 2nd and 98th centiles.
Results
The resulting z-score charts look pl Back-transforming the mean ±2 SD lines provided gene-specific median, 2nd and 98th centiles.
Results
The resulting z-score charts look plausible on several counts. Only *KMT2A* shows a (rising) age
in median height, wh In median height, while BMI and weight increase in several genes, possibly reflecting population
trends. Apart from *SATB2* and *DDX3X*, the gene-specific medians are all below the reference (range
0.1th centile for heig trends. Apart from *SATB2* and *DDX3X*, the gene-specific medians are all below the reference (rand 0.1th centile for height *KMT2A* to 36th centile for BMI *ANKRD11*). Median OFC shows no age trence with medians rangi 0.1th centile for height *KMT2A* to 36th centile for BMI *ANKRD11*). Median OFC shows no age trend,
0.1th centile for height *KMT2A* to 36th centile, and *ASXL3* lowest, on the 3rd centile. There are no sex
diffe

Conclusions

0.1" centile for height *KMT2A* to 36" centile for BMI *ANKRD11*). Median OFC shows no age trend,
with medians ranging from 10th-30th centile, and ASX*L3* lowest, on the 3rd centile. There are no sex
differences in 1 with medians ranging from 10"-30" centile, and ASXL3 lowest, on the 3" centile. There are no sex
differences in 19/24 cases.
Conclusions
Our LMSz method produces gene-specific growth charts for rare diseases, an essential Conclusions
Our LMSz method produce:
for paediatric care. We plar
relevant genes. for paediatric care. We plan to automate it within the DECIPHER platform, enabling availability for
relevant genes.
 for parameters it within the DECIPHER plan to automate it within the DECIPHER platform, enabling availability for all
relevant genes.
3 relevant genes.

Introduction

the trajectory against reference centiles (1). Many children with genetic disorders plot on an
extreme centile—in itself a clue to a possible underlying genetic diagnosis (2). Once a genetic
diagnosis is made, plotting on the trajectory against reference centure (1). The trajectory genetic diagnosis (2). Once a genetic
extreme centile—in itself a clue to a possible underlying genetic diagnosis (2). Once a genetic
diagnosis is made, plotting extreme centre and a cluster a positive and any be misleading. It may, for example, sugge
child is of short stature and underweight where they are growing normally for their genetic di
leading to clinical/parental anxiety diagnosis is of short stature and underweight where they are growing normally for their genetic disord
leading to clinical/parental anxiety resulting in unnecessary investigation and
unwarranted/ineffective intervention. C leading to clinical/parental anxiety resulting in unnecessary investigation and
unwarranted/ineffective intervention. Conversely, abnormal growth may be incorrectly ascribed to
the underlying genetic disorder, and other ca unwarranted/ineffective intervention. Conversely, abnormal growth may be in
the underlying genetic disorder, and other causes left untreated. Gene-specific
therefore important in paediatric care. However, due to the small underlying genetic disorder, and other causes left untreated. Gene-specific growth charts are
therefore important in paediatric care. However, due to the small numbers of affected individuals,
few gene-specific growth char

therefore important in paediatric care. However, due to the small numbers of affected individuals
few gene-specific growth charts are available (3-5).
This proof-of-principle study tests a new method for constructing growt therefore, permany permany and substitute in parameteristic care in a measurement in parameters.
This proof-of-principle study tests a new method for constructing growth charts in rare disorders
based on small datasets, wh

Methods

Fhis proof-of-principle study tests a new method fo
based on small datasets, which we call the LMSz me
method (9) on the z-score scale.
Methods
Growth data were collated from the Deciphering De
study (7) datasets, both fro The proof-principle study (9) on the z-score scale.
The method (9) on the z-score scale.
Methods
Growth data were collated from the Deciphering Developmental Disorders (DDD) (6) and GenROC
Study (7) datasets, both from the based (9) on the z-score scale.

Methods

Growth data were collated from the Deciphering Developmental Disorders (DDD) (6) and GenROC

study (7) datasets, both from the UK. For proof-of-principle we selected ANKRD11, a fre methods
Growth data were collated from
study (7) datasets, both from the
diagnosed gene in the DDD study
with pathogenic or likely pathoge
diagnoses. We extracted data on
circumference (OFC), and longitu study (7) datasets, both from the UK. For proof-of-principle we selected *ANKRD11*, a frequently
diagnosed gene in the DDD study with a well described phenotype. We included only participants
with pathogenic or likely path study (7) datasets, both from the OK. For proof-of-principle we selected ANKRD11, a frequently
diagnosed gene in the DDD study with a well described phenotype. We included only participani
with pathogenic or likely pathoge Multipality of the Multipality parameters and parameters principality parameters any parameters of the parameters diagnoses. We extracted data on sex, gestational age, birth weight, birth occipitofrontal circumference (OFC with pathogenic or likely pathogenic ANKRD11 variants, excluding any with composite genetic
diagnoses. We extracted data on sex, gestational age, birth weight, birth occipitofrontal
circumference (OFC), and longitudinal me

diagnoses of height, weight and OFC and the associat measurement.
at measurement.
For ANKRD11, we undertook a second phase of analysis including European datasets (Sp
Denmark and The Netherlands). We also extended the anal at measurement.
For ANKRD11, we undertook a second phase of analysis including European datasets (Spain,
Denmark and The Netherlands). We also extended the analysis to five other genes: ARID1B, ASXL3,
DDX3X, KMT2A and SATB at measurement
For ANKRD11, we
Denmark and The
DDX3X, KMT2A ar For ANKRD11, we undertook a second phase of analysis including European datasets (Spain,
Denmark and The Netherlands). We also extended the analysis to five other genes: ARID1B, .
DDX3X, KMT2A and SATB2. ASXL3 growth data DDX3X, KMT2A and SATB2. ASXL3 growth data were collated from the DDD study and the ASXL3
DDX3X, KMT2A and SATB2. ASXL3 growth data were collated from the DDD study and the ASXL3 DDX3X, KMT2A and SATB2. ASXL3 growth data were collated from the DDD study and the ASXL3

Siris Syndrome heterogeneous gene group growth charts (4).
Siris Syndrome heterogeneous gene group growth charts (4).
We also analysed data from the Mowat Wilson Syndrome (MWS) Growth Chart Consortium and
compared our res Signal We also analysed data from the Mowat Wilson Syndrome (M'

compared our results with those obtained by fitting the raw of

Supplement 1 for details). They were also compared to the co

by the MWS Consortium (3).

Age compared our results with those obtained by fitting the raw data using the LMS method (see
Supplement 1 for details). They were also compared to the corresponding centile charts publishe
by the MWS Consortium (3).
Age was Supplement 1 for details). They were also compared to the corresponding centile charts publi
by the MWS Consortium (3).
Age was adjusted for gestation using the formula *adjusted age* = $age + (gestation - 40)$
7/365.25. Body mass in

fusied dye – dye τ (yestation – 40) \times

Statistical analysis

By the MWS Consortium (3).

Age was adjusted for gestation using the formula adjusted $age = age + (gestation - 40) \times 7/365.25$. Body mass index (BMI) was calculated as *weight/height*² in units of kg/m^2 .

Statistical analysis

Growth By measure external (4).
Age was adjusted for gestatic
7/365.25. Body mass index (
Statistical analysis
Growth centile charts are cor
a special case of the family of
(9). Using anthropometry fro
smooth curves in age. M is $7/365.25$. Body mass index (BMI) was calculated as
Statistical analysis
Growth centile charts are conventionally constructed
a special case of the family of Generalized Additive N
(9). Using anthropometry from a referenc *7/365.25. Body mass index (BMI) was calculated as weight/height* - in units or $\kappa g/m$ -.

Statistical analysis

Growth centile charts are conventionally constructed using the LMS (lambda-mu-sigma) n

a special case of th a special case of the family of Generalized Additive Models for Location Scale and Shape (GAMLSS)
(9). Using anthropometry from a reference sample of children the LMS method estimates three
smooth curves in age. M is the m (9). Using anthropometry from a reference sample of children the LMS method estimates three
smooth curves in age. M is the median or 50th centile curve; the S curve is the coefficient of variatio
or fractional standard (9). Using antia-parties) with a reference sample of children the last in the decefficient of varia
or fractional standard deviation (SD), and the L curve is a measure of skewness. Any required ce
curve can be constructed

smooth curves in age. M is the median or 50"" centile curve; the S curve is the coefficient of variation
or fractional standard deviation (SD), and the L curve is a measure of skewness. Any required centile
curve can be co or fractional standard deviation (SD), and the Latitude deviation of summations), is a mode converted to exact z-scores.
The LMS method is most effective with large datasets, ideally many thousands of individuals (10).
The Experience constrained from the converted for exact 2-scores.
The LMS method is most effective with large datasets, ideally many thousands of individuals (10).
The issue with charts for rare diseases is the inherent scarci The LMS method is most effecti
The issue with charts for rare di
be based solely on the available
how children grow, as containe
reference can then be adjusted
This is analogous to a Bayesian a
combined with rare disease d The issue with charts for rare diseases is the inherent scarcity of data. In this case, the chart cannot be based solely on the available data, it must "borrow strength" from previous knowledge about how children grow, as The issue with the interact in the interact of the interact of the interact of the based solely on the available data, it must "borrow strength" from previous knowledge about
how children grow, as contained in an existing be based in an existing or "baseline" growth reference. This baseline
reference can then be adjusted using the rare disease data to obtain a reference for the rare dise
This is analogous to a Bayesian analysis where prior reference can then be adjusted using the rare disease data to obtain a reference for the rare
This is analogous to a Bayesian analysis where prior information (the baseline reference) is
combined with rare disease data to This is analogous to a Bayesian analysis where prior information (the baseline reference) is
combined with rare disease data to give a posterior or updated reference relevant to the rare
disease.
5 combined with rare disease data to give a posterior or updated reference relevant to the radisease. disease data to give a posterior or updated rare disease data to give a posterior or updated reference relevant to the rare disease α

Free is the British 1990 reference (UK90) (11), the official UK chart for birth and age 4-20 and the reference used by the DECIPHER database. Switching to the z-score scale creates centile curves which are horizontal lines reference used by the DECIPHER database. Switching to the z-score scale creates centile curves
which are horizontal lines, as each centile curve connects points with the same z-score; they are a
the same for both sexes. Fo which are horizontal lines, as each centile curve connects points with the same z-score; they are
the same for both sexes. For representative children the z-transformed centiles are Normally
distributed with mean (and medi the same for both sexes. For representative children the z-transformed centiles are Normally
distributed with mean (and median) 0 and SD 1 at all ages. Both mean and SD can then be viewed as
straight lines plotted against

distributed with mean (and median) 0 and SD 1 at all ages. Both mean and SD can then be vies
traight lines plotted against age. Other centile curves are lines above or below the median, s
according to the corresponding z-s straight lines plotted against age. Other centile curves are lines above or below the median, spaced
according to the corresponding z-score (e.g. the 97.7th centile is at z = 2, i.e. 2 SDs above the mean).
The next stage strategy in the corresponding z-score (e.g. the 97.7th centile is at z = 2, i.e. 2 SDs above the mean).
The next stage is to assume that the rare disease median curve is also a straight line but estimated
from the rare d according to the corresponding z-score (e.g. the 97.7" centile is at z = 2, i.e. 2 SDs above the mean).
The next stage is to assume that the rare disease median curve is also a straight line but estimated
from the rare dis From the rare disease data as the linear regression of UK90 z-score on age and sex. The SD curve is
estimated in the same way from the linear regression of the z-score SD on age and sex. GAMLSS is
used to estimate the mean estimated in the same way from the linear regression of the z-score SD on age and sex. GAMLSS is
used to estimate the mean and SD regression lines simultaneously, using the Normal or NO
distribution family. The two lines a estimate the mean and SD regression lines simultaneously, using the Normal or NO
distribution family. The two lines are analogous to the M and S curves of the LMS method, but on
the z-score scale. Note that there is no equ used to the M and S curves of the LMS method, between scale. Note that there is no equivalent L line as the normal distribution is by defined skew, and the conversion from measurement to z-score is assumed to have removed the z-score scale. Note that there is no equivalent L line as the normal distribution is by definition
not skew, and the conversion from measurement to z-score is assumed to have removed any
skewness.
The GAMLSS regressio

the z-score is assumed to have removed any
not skew, and the conversion from measurement to z-score is assumed to have removed any
skewness.
The GAMLSS regression equation for the M line is made up of four terms: intercep skewness.
The GAMLSS regression equation for the M line is made up of four terms: intercept, age trend
sex difference and age by sex interaction (i.e. the age trend differing between the sexes), and
coefficient for each t The GAML
sex differe
coefficient
M
But if one
dropped fr
M Sex difference and age by sex interaction (i.e. the age trend differing between the sexes), and the coefficient for each term is estimated:
 $M = a_M + b_M \times age + c_M \times sex + d_M \times age \times sex$

But if one or more of the fitted coefficients is sm

 $\mu_M + \nu_M \wedge \mu$ ye + $\mu_M \wedge$ sex + μ

coefficient for each term is estimated:
 $M = a_M + b_M \times age + c_M \times sex + d_M \times age \times sex$

But if one or more of the fitted coefficients is small enough, the corresponding term (s) can be

dropped from the regression, leading to a set of five $M = a_M + b_M \times age + c_M \times s$
But if one or more of the fitted coeffici
dropped from the regression, leading t
 $M = a_M + b_M \times age + c_M \times s$
 $M = a_M + b_M \times age$ But if one or
dropped from
 $M =$
 $M =$ $M \wedge uye \wedge sex$

small enough, the

progressive progressive dropped from the regression, leading to a set of five progressively simpler equations:
 $M = a_M + b_M \times age + c_M \times sex$
 $M = a_M + b_M \times age$

$$
M = a_M + b_M \times age + c_M \times sex
$$

$$
M = a_M + b_M \times age
$$

$$
M = a_M + b_M \times age
$$

$$
M = a_M + b_M \times age
$$

$$
M = a_M + c_M \times sex
$$

$$
M = a_M
$$

$$
M = 0
$$

 $M =$
 $M =$
The SD or S li
interaction to
 log (

$$
M = a_M
$$

\n
$$
M = 0
$$

\nThe SD or S line is modelled in the same way, except it is log transformed to ensure positivity. The
\ninteraction term is also omitted for simplicity.
\n
$$
log(S) = a_S + b_S \times age + c_S \times sex
$$

\n
$$
log(S) = a_S + b_S \times age
$$

\n
$$
log(S) = a_S + c_S \times sex
$$

\n
$$
log(S) = a_S
$$

\n
$$
log(S) = 0
$$

\nNote that the final equations in these lists are null models where the coefficient is 0. They
\ncorrespond to M = 0 and S = 1, the mean and SD for the baseline UK90 reference, which are also th
\ndefault values for the GAMLSS NO distribution. GAMLSS estimates the M line and S line
\nsimultaneously, and there are 6 M line equations and 5 S line equations, so there are 6 × 5 = 30

 $log(S) =$
 $log(S) =$
 $log(S) =$
 $log(S) =$
Note that the final
correspond to M =
default values for $log(S) =$
 $log(S) =$
 $log(S) =$
Note that the final
correspond to M =
default values for
simultaneously, ar
possible model co a_S
 a_S
 0
 0

equations in the damental parameters
 α and $S = 1$, the damental parameters
 α there are 6 M
 α there are 6 M
 α there are 6 M $log(S) =$
 $log(S) =$
Note that the final
correspond to M =
default values for
simultaneously, ar
possible model co
goodness of fit usi $\begin{matrix} a_s \end{matrix}$
0 equ $\begin{matrix} 0 & a \end{matrix}$
he dtl
nbi Note that the final
correspond to M =
default values for
simultaneously, ar
possible model co
goodness of fit usi
criterion (BIC), the
linear on the log s ove Ohedink ng orali correspond to M = 0 and S = 1, the mean and SD for the baseline UK90 reference, which are default values for the GAMLSS NO distribution. GAMLSS estimates the M line and S line simultaneously, and there are 6 M line equati default values for the GAMLSS NO distribution. GAMLSS estimates the M line and S line
simultaneously, and there are 6 M line equations and 5 S line equations, so there are $6 \times 5 = 30$
possible model combinations. All 30 a simultaneously, and there are 6 M line equations and 5 S line equations, so there are 6 :
possible model combinations. All 30 are fitted to the data in turn, and they are compare
goodness of fit using either the Akaike inf simulaneously, and there are 6 M line equations and 3 S line equations, so there are 6 \times 3 –
possible model combinations. All 30 are fitted to the data in turn, and they are compared for
goodness of fit using either th e is
30
add
er
bet
er. F goodness of fit using either the Akaike information criterion (AIC) or the Bayesian information
criterion (BIC), the optimal model being the one with the lowest AIC or BIC. Note that the S lir
linear on the log scale, but criterion (BIC), the optimal model being the one with the lowest AIC or BIC. Note that the S line
linear on the log scale, but backtransformed it is slightly curved if the b_S coefficient is non-zerc
Lack of fit is measu linear on the log scale, but backtransformed it is slightly curved if the b_S coefficient is non-zero.
Lack of fit is measured by the deviance, or -2 log likelihood. The AIC penalises the deviance by add
two units for ea linear on the log scale, but backtransformed it is slightly curved if the *bg* coefficient is non-zero.
Lack of fit is measured by the deviance, or -2 log likelihood. The AIC penalises the deviance by at
two units for each Experience in the deviation of the model, while the BIC uses a larger penalty of log *n* units per
coefficient, where *n* is the sample size. This penalises more complex models, which tend to fit better.
For a model where

two units for each coefficient in the model, while the BIC uses a larger penalty of log n units per
coefficient, where n is the sample size. This penalises more complex models, which tend to fit be
For a model where $n = 7$ For a model where $n = 7$ the AIC and BIC are the same, but for larger *n* the BIC penalty is larger. For example, the most numerous measurement is weight in *ANKRD11* where there are 488 points and 7 For a model where $n =$
example, the most num 7 the AIC and BIC are the same, but for larger *n* the BIC penalty is larger. For
erous measurement is weight in *ANKRD11* where there are 488 points and
7 example, the most numerous measurement is weight in ANKRD11 where there are 488 points and

 $log(4889) \approx 6$, so here the BIC penalty is three times the AIC penalty. For this reason, the BIC
models are either the same as or simpler than the AIC models.
For larger samples the set of models for the M line can be exte For larger samples the set of models for the M line can be exte
in age as an alternative to a straight line. This increases the nur
to 9, and the total of combined M+S models from 30 to 45.
Once the optimal model has been

For a sample in a straight line. This increases the number of possible fitted M lines from 6

to 9, and the total of combined M+S models from 30 to 45.

Once the optimal model has been identified, it is used to predict the in age as an alternative to a straight line. In anti-article in a straight line in the number of possible incre
Once the optimal model has been identified, it is used to predict the required centile lines on the
UK90 z-sco Once the optimal model has been identified, it is used to pr
UK90 z-score scale. They are then backtransformed to the n
reference, giving a set of centile curves appropriate for the
Error-checking
To detect possible data o

Error-checking

UK90 z-score scale. They are then backtransformed to the measurement scale using the UK90
reference, giving a set of centile curves appropriate for the rare disease.
Error-checking
To detect possible data outliers, the dat reference, giving a set of centile curves appropriate for the rare disease.
Error-checking
To detect possible data outliers, the data in the form of UK90 z-scores were centred and scale
gene-specific z-scores, and z-scores Frror-checking
To detect possible data outliers, the data in the form of UK90 z-scores we
gene-specific z-scores, and z-scores exceeding 3.5 in absolute value were
In this way several data errors were identified and either

Specificity

The sepecific z-scores, and z-scores exceeding 3.5 in absolute value were excluded from the analys
In this way several data errors were identified and either corrected or excluded.
For simplicity the analysis treats any re In this way several data errors were identified and either corrected or excluded.
For simplicity the analysis treats any repeated measures in the data as independent.
Specificity
To test the specificity of the method, z-s For simplicity the analysis treats any repeated measures in the data as independent
Specificity
To test the specificity of the method, z-scores of measurements with ANKRD11 (
numerous) were replaced by z-scores randomly s Specificity

To test the specificity of the method, z-scores of measurements with ANKRD11 (bein

numerous) were replaced by z-scores randomly sampled from a standard Normal dis

mean 0 and SD 1. In this way simulated z-sc To test the specificity of the method, 2-scores of measurements with ANKRD11 (being the most
numerous) were replaced by 2-scores randomly sampled from a standard Normal distribution, w
mean 0 and SD 1. In this way simulat mean 0 and SD 1. In this way simulated z-score data were generated with the age and sex structure
of the underlying data, where the optimal model ought to be $M = 0$ and $log(S) = 0$, correspondin
to the baseline UK90 referenc of the underlying data, where the optimal model ought to be $M = 0$ and $log(S) = 0$, correspondint to the baseline UK90 reference. The model was then fitted repeatedly to newly sampled data, and the optimal model each time no of the underlying data, where the optimal model ought to be $m =$
to the baseline UK90 reference. The model was then fitted repeate
the optimal model each time noted, 100 times each for height (n =
(n = 343) and OFC (n = 2 $\frac{dy}{dy}$ to newly sam
355 points), weig β pled data, and
ht (n = 488), BMI the optimal model each time noted, 100 times each for height (n = 355 points), weight (n = 488), BI
(n = 343) and OFC (n = 231). the optimal model each time noted, 100 times each for height (n = 12 points), weight (n = 12 points), weight (n = 488), BMI (n = 343), BMI (n = 231). $(n - 34)$ and $(n - 23)$.

Results

The age range was restricted to 0-18 years. Linked ethnicity data were not recorded, but we assume
the vast majority of cases were Caucasian based on the sources.
Table 1: counts of the number of subjects and data by gene, The vast majority of cases were Caucasian based on the sources.
Table 1: counts of the number of subjects and data by gene, country, sex and measurement. *NatS =
Natural History Study (International)

the vast majority of cases were calculated on the sources.
Table 1: counts of the number of subjects and data by gene, cour
Natural History Study (International) $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Table 1: counts of the number of subjects and data by gene, country, sex and measurement. *Natural History Study (International) Natural History Study (International)

Z-score centile charts and growth charts

 $\frac{1}{2}$ d c r The results are presented as a rectangular array of centile charts by gene (ANKRD11, ARRD11, ASL3), $DDX3X$, $KMT2A$ and $SATB2$ for height, weight, BMI, and OFC (Figure 1). Three centiles are displayed, corresponding to z-sc Corresponding to z-scores -2, 0 and 2, i.e. the 2.3rd, 50th and 97.7th centiles. For simplicity they are referred to here as the 2nd, 50th and 98th centiles.
The referred to here as the 2nd, 50th and 98th corresponding to z-scores -2, 0 and 2, i.e. the 2.3rd, 50th and 97.7th centiles. For simplicity they are referred to here as the 2nd, 50th and 98th centiles. referred to here as the 2nd, 50th and 98th centiles.

In the data from all four countries, while Supplement 2 discusses the appreciable inter-country

differences in ANKRD11 height.

Figure 1 penalises the plots using the BIC, while Supplementary Figure 2 shows the same plots the data from all four countries, while Supplement 2 discusses the appreciable inter-country
differences in *ANKRD11* height.
Figure 1 penalises the plots using the BIC, while Supplementary Figure 2 shows the same plots
pe Figure 1 penalises the plots using the BIC, while Supplementary Figure 2 shows the same plot
penalised with the AIC. As expected, the two sets of centiles are broadly similar, though 13 o
plots in SF2 show a sex difference Figure 1 penalises the plots usin
penalised with the AIC. As expected
plots in SF2 show a sex difference
the BIC plots of Figure 1 with the
spline curve, and in all cases the
The specificity of the method waterspectively 9 Figure 1 penalised with the AIC. As expected, the two sets of centiles are broadly similar, though 13 of t
plots in SF2 show a sex difference as against five in Figure 1. Similarly, Supplementary Figure 3
the BIC plots of

plots in SF2 show a sex difference as against five in Figure 1. Similarly, Supplementary Figure 3 shows
the BIC plots of Figure 1 with the option of a spline curve for the M line. Five of the 24 facets select a
spline curv plant the BIC plots of Figure 1 with the option of a spline curve for the M line. Five of the 24 facets select a
spline curve, and in all cases they show a dip in early life compared to later childhood.
The specificity of spline curve, and in all cases they show a dip in early life compared to later childhood.
The specificity of the method was assessed with the *ANKRD11* data. Using the BIC models,
respectively 98, 97, 97 and 97 times out o The specificity of the method was assessed with the *ANKRD11* data. Using the BIC modespectively 98, 97, 97 and 97 times out of 100 the selected model was the null model, appropriate for the baseline growth reference. This The specificity of the method was assessed with the ANKRD11 data. Using the BIC models, i.e.
The specificity of the baseline growth reference. This corresponds to an average specificity of
over 97%, i.e. the percentage of respectively appropriate for the baseline growth reference. This corresponds to an average specificity of over 97%, i.e. the percentage of times the procedure correctly chooses the baseline referent data from non-syndromic approximate for the percentage of times the procedure correctly chooses the baseline reference for data from non-syndromic children. For comparison, repeating the exercise using the AIC models gave a much lower specificity

data from non-syndromic children. For comparison, repeating the exercise using the AIC models
gave a much lower specificity of 53%.
Overall, the differences between Figure 1 and Supplementary Figures 2 and 3 are small, and

data from non-syntantic and the exercy of 53%.

Overall, the differences between Figure 1 and Supplementary Figures 2 and 3 are small, and give

their lower specificity there is no obvious reason to prefer the more complic Overall, the differences between Figure
their lower specificity there is no obvior
the BIC-based centiles, so the focus fro
Supplementary Table 1 shows the reg
Figure 1. Except for weight and BMI for
coefficients, some of Their lower specificity there is no obvious reason to prefer the more complicated AIC centiles over
the BIC-based centiles, so the focus from here on is the BIC centiles of Figure 1.
Supplementary Table 1 shows the regress the BIC-based centiles, so the focus from here on is the BIC centiles of Figure 1.
Supplementary Table 1 shows the regression coefficients and standard errors for the models in
Figure 1. Except for weight and BMI for SATB2 Supplementary Table 1 shows the regression coefficients and standard errors for
Figure 1. Except for weight and BMI for *SATB2*, all the models include one or mc
coefficients, some of them highly so, despite the small samp Figure 1. Except for weight and BMI for SATB2, all the models include one or more significant
coefficients, some of them highly so, despite the small sample sizes involved. Conversely no mo
includes an age by sex interacti righted 2. Except for weight and BMI for SATB2, all the models include one or more significant
coefficients, some of them highly so, despite the small sample sizes involved. Conversely no n
includes an age by sex interacti includes an age by sex interaction even though it was an option. Altogether there are 53 regression
coefficients fitted, which with 24 distinct gene-measure models correspond to just over two
coefficients per model. Supple includes an age by sex interaction even though it was an operation. Some more are survey contributed to the pro
coefficients fitted, which with 24 distinct gene-measure models correspond to just over two
coefficients per m coefficients per model. Supplementary Table 2 shows the 74 corresponding coefficients for t
models. coefficients per models.
models.
3

Gene-specific charts

Figures 2 and 3 show the 2^{no}, 50^{oo} and 98^{oo} measurement centiles obtained by back-transforming the
2-score centiles of Figure 1, for females and males respectively, as a rectangular 4 x 6 plot by
measurement and gene z-score centiles of Figure 2, the Corresponding UK90 centiles shown as dashed lines.

Gene-specific charts

In 19 of the 24 panels in Figure 1, the centiles for males and females are the same, including

In 19 of the 24 pa Gene-specific charts
In 19 of the 24 panels in Figure 1, the centiles for males and females are the same, incl
measurements for ANKRD11, ARID1B and KMT2A. The sex differences where they app
generally small. None of the gen measurements for ANKRD11, ARID1B and KMT2A. The sex differences where they appear are
generally small. None of the genes show an age trend in median OFC, and only KMT2A shows a
(rising) age trend in median height. But for measurements for AnkhD11, AMD1D and KMT2A. The sex differences where they appear are
generally small. None of the genes show an age trend in median OFC, and only KMT2A shows
(rising trends, by up to +2 SDs from birth to ag (rising) age trend in median height. But for weight and BMI two and four genes respectively showing trends, by up to +2 SDs from birth to age 18. This may reflect increasing adiposity with age the UK90 reference correspon rising trends, by up to +2 SDs from birth to age 18. This may reflect increasing adiposity with age, it
he UK90 reference corresponds to child adiposity as seen in 1990, and obesity prevalence increas
with age (12). In 20 the UK90 reference corresponds to child adiposity as seen in 1990, and obesity prevalence increases
with age (12). In 20 of the panels the three centiles are parallel lines, indicating that the variability is
constant acr with age (12). In 20 of the panels the three centiles are parallel lines, indicating that the variability is
constant across age. The most striking exception is for weight with *DDX3X*, where the variability is
dramatical with age (12). In 20 of the particle in the three centiles are parallel lines, indicating that the variability is
constant across age. The most striking exception is for weight with *DDX3X*, where the variability is
dramat

50th centile

constant across age. The most striking exception is for weight with DDX3X, where the variability is
dramatically greater at age 18 than at birth.
Supplementary Table 3 shows the UK90 centiles at birth corresponding to the Mammaram, greater at age 18 than at birth.
Supplementary Table 3 shows the UK90 cent
centiles for each measurement, averaged ac
50th centile
The median for SATB2 (height, weight and B
UK90 median. However for all other m Supplementary Table 3 shows the UK90 centiles at birth corresponding to the 2nd, 50th and 98th
centiles for each measurement, averaged across sex, by gene.
50th centile
The median for SATB2 (height, weight and BMI) 50th centile
The median for SATB2 (height, weight and BMI) and for DDX3X
UK90 median. However for all other measurements and genes
reference, ranging from 0.08th (height KMT2A) to 28th (BMI AN
98th centile
The 98th

98th centile

UK90 median. However for all other measurements and genes the median is much lower than
reference, ranging from 0.08th (height KMT2A) to 28th (BMI ANKRD11) UK90 centile.
98th centile
The 98th centiles are nearly all a untain. However for all other measurements and general measurements and denote measurements and the measurement
198th centile
198th centile
The 98th centiles are nearly all above the UK90 90th centile. However 98th c reference, ranging from 0.08th (height KMT2A) to 28" (BMI ANKRD11) UK90 centile.
98th centile
The 98th centiles are nearly all above the UK90 90th centile. However 98th centile for h
is on the 44th UK90 centile. Ve The 98"" centiles are nearly all above the UK90 90"" centile. However 98"" centile for height in KMT2A
is on the 44th UK90 centile. Very few *KMT2A* cases achieve even median height on the UK90 chart.
Other relatively l is on the 44" UK90 centile. Very few *KMT2A* cases achieve even median height on the UK90 chart.
Other relatively low 98th centiles are for height in *ASXL3* (79th) and weight in *DDX3X* (76th). Other relatively low 98" centiles are for height in ASXL3 (79") and weight in *DDX3X* (76").

States are for height in ASXL3 (79") and weight in *DDX3X* (76").

The control of the control of the control of the control of t

2nd centile as z-score

Mowat Wilson Syndrome (MWS) – ZEB2

The 2nd centiles at birth are shown as UK90 z-scores because many are far below the 1^{ch} UK90
centile. The most extreme cases are *KMT2A* (-6 SD) and *ARID1B* (-5 SD), while all the 2nd centil
on or below the UK90 2ⁿ centile. The most extreme cases are *KMT2A* (-6 SD) and *ARID1B* (-5 SD), while all the 2nd centiles are
on or below the UK90 2nd centile.
Mowat Wilson Syndrome (MWS) – *ZEB2*
Given the large sample size, we estimated on or below the UK90 2nd centile.
Mowat Wilson Syndrome (MWS
Given the large sample size, we es
the z-scores (Figure 4A), and the L
scale (Figure 4A) all four measurer
median close to the UK90 2nd cent
rates for weight the z-scores (Figure 4A), and the LMS method as applied to the raw data (Figure 4B). On the z-score
scale (Figure 4A) all four measurement medians are below the UK90 median at birth, with the OFC
median close to the UK90 2 scale (Figure 4A) all four measurement medians are below the UK90 median at birth, with the OFC
median close to the UK90 2nd centile. The medians all fall further with increasing age, and at faster
rates for weight and O scale (Figure 4B) consistent by sext, below the UK90 2^{nd} centile. The medians all fall further with increasing age, and at faster rates for weight and OFC among the females. Figure 4B compares the 2^{nd} , 50th and 98 median close to the UK90 2^{no} centile. The medians all fall further with increasing age, and at faster
rates for weight and OFC among the females. Figure 4B compares the 2nd, 50th and 98th centiles by
the two method rates for weight and OFC among the females. Figure 4B compares the 2nd, 50^{dd} and 98^d centiles by the two methods, for the sexes separately, where the solid curves correspond to the LMS method while the dashed curves while the dashed curves are for the LMSz method. For height, weight and BMI there is reasonable
agreement, though the BMI 98th centiles agree less well. The height and weight centiles are
consistent by sex, being similar agreement, though the BMI 98th centiles agree less well. The height and weight centiles are
consistent by sex, being similar until puberty when the males become taller and heavier. However,
unlike the LMSz centiles, the agreement, though the BMI 98" centiles agree less well. The height and weight centiles are
consistent by sex, being similar until puberty when the males become taller and heavier. Ho
unlike the LMSz centiles, the LMS centi unlike the LMSz centiles, the LMS centiles continue rising after puberty despite growth ending,
because there are too few data at older ages to pull them down. OFC performs poorly, with all thre
LMS-based centiles rising a

Discussion

because there are too few data at older ages to pull them down. OFC performs poorly, with all LMS-based centiles rising and falling at different ages, while the LMSz centiles rise linearly for n
but fall in later childhood EMS-based centiles rising and falling at different ages, while the LMSz centiles rise linearly for males
but fall in later childhood for females.
Discussion
We have developed a method for producing gene-specific z-score ce but fall in later childhood for females.
Discussion
We have developed a method for producing gene-specific z-score centile charts and growth charts
based on small datasets in rare disease. The resulting charts are unbiased Discussion
We have developed a method for proc
based on small datasets in rare disease
constructed, cosmetically they look pla
centiles almost always exactly match t
specificity is as high as 97%. we hased on small datasets in rare disease. The resulting charts are unbiased due to the way they are
constructed, cosmetically they look plausible, they agree with other published literature, and the
centiles almost alway based on small datasets in results in results in rare distance to the results in rare constructed, cosmetically they look plausible, they agree with other published literature, and the centiles almost always exactly match centiles almost always exactly match the baseline reference when given random z scores ie. the
specificity is as high as 97%. specificity is as high as 97%.
Specificity is as high as 97%. s_p is as high as $\frac{1}{2}$ is as $\frac{1}{2}$

Sample size

outer centiles (10). This does not work for the small samples available in rare disease cohorts. The
LMSz method instead considers the universe of all possible simple models, such as all combinations
of sex and linear age LMSz method instead considers the universe of all possible simple models, such as all combination
of sex and linear age for the mean and SD, and selects the best model, penalised for complexity. Th
works because the unive

Fitting of centiles

of sex and linear age for the mean and SD, and selects the best model, penalised for complexity. This
works because the universe of all possible linear models is small, between 30 and 45 alternative
models, and the model s or shown and any of section the mean and set in the means and spiritual in the means of sections and the model selection does not depend on statistical significance.
Fitting of centiles
Due to the larger penalty per extra works because the model selection does not depend on statistical significance.
Fitting of centiles
Due to the larger penalty per extra df, the BIC models ought to be simpler than the AIC models, a
indeed they are—the BIC m Fitting of centiles
Due to the larger penalty per extra df, the BIC models ought to be simpler the
indeed they are—the BIC models underlying the 24 facets in Figure 1 involve
coefficients per facet. In contrast the corresp Indeed they are —the BIC models underlying the 24 facets in Figure 1 involve 53 coefficients, i.e. ~2
coefficients per facet. In contrast the corresponding AIC models in Supplementary Figure 2 involve
74 coefficients, aver coefficients per facet. In contrast enteringing and a therm of and anti-time are coefficients per facet. In contrast the corresponding AIC models in Sipplementary Figure 2 involve
74 coefficients, averaging over 3 per face The coefficients, averaging over 3 per facet. Despite this the centiles in Figure 1 and Supplementary
Figure 2 are not materially different, probably because the coefficients are all small. The specificity
calculation show Figure 2 are not materially different, probably because the coefficients are all small. The specificity
calculation showed that the AIC models had only 53% specificity, compared to 97% for the BIC
models, for selecting the Figure 2 are not materially antitially present, present to compared to 97% for the BIC
calculation showed that the AIC models had only 53% specificity, compared to 97% for the BIC
models, for selecting the correct null mod material method into the AIC model when the data were random Normal. This shows the
models, for selecting the correct null model when the data were random Normal. This shows the
AIC tends to choose over-complex models, whi AIC tends to choose over-complex models, which in turn supports the use of the BIC models. Note
that the sensitivity of the method cannot be tested in the same way as the specificity. It involves
knowing in advance the cen

Z-scores versus growth centiles

Matthe sensitivity of the method cannot be tested in the same way as the specificity. It involves
knowing in advance the centiles appropriate for the particular genotype, which are by definition
unknown.
Z-scores versus gr that the sensitivity of the method cannot provide the method can be the method can be the method can
that that the method can unknown.
The z-score sensitive plots show clearly the differences in growth pattern for a partic Express versus growth centiles
The z-score centile plots show clearly the differences in growth pattern for a particular gene
compared to UK90. The z-score plots compactly show growth trends over time, and sex difference
w Z-scores v
The z-score
compared
where pre:
growth on The z-score plots show growth trends over time, and sex differences or parameters of the z-score plots compactly show growth trends over time, and sex differences present. However, in clinical practice health professionals where present. However, in clinical practice health professionals measure and plot their patients'
growth on centile charts as part of standard clinical care. For this reason, it is important to provide
1 growth on centile charts as part of standard clinical care. For this reason, it is important to provide
growth on centile charts as part of standard clinical care. For this reason, it is important to provide growth on centile charts as part of standard clinical care. For this reason, it is important to provide

Comparison with the literature and published charts

the back-transformed centrical comparison with the literature and published charts
There is a published centile chart for Coffin-Siris Syndrome (CSS; OMIM #135900), caused by varian
in BAF complex genes (4). Direct compar Comparison with the literature and
There is a published centile chart for
in BAF complex genes (4). Direct com
heterogeneity and differing endpoint
sex difference for weight, but the 50^{ti}
consistent.
We were able to anal In BAF complex genes (4). Direct comparison with $ARID1B$ is not possible given the greater genetic
heterogeneity and differing endpoints (height to age 10; OFC to 36 months). In fact, we did not see a
sex difference for we

in BAF complex genes (4). Direct comparison with ARB210 is not possible given the greater genetic
heterogeneity and differing endpoints (height to age 10; OFC to 36 months). In fact, we did not see
sex difference for weigh sex difference for weight, but the 50th centile for height at age 10, and weight and OFC were
consistent.
We were able to analyse the MWS dataset in two ways: LMSz as described here and the
conventional LMS method appli sex difference for weight, but the 50th centile for height at age 10, and weight and OFC were
consistent.
We were able to analyse the MWS dataset in two ways: LMSz as described here and the
conventional LMS method applie We were ak
convention:
though the
ages. For BI
probably be
but the deg
affect the u conventional LMS method applied to the raw data. The two sets of centiles are reasonab
though the z-score-based centiles look more convincing for height and weight, particular
ages. For BMI, the 2nd and 50th centiles a computed a LMS are the range of the range applies the range applies are more discordant. This is probably because the 2-score conversion applies the UK90 ske ages. For BMI, the 2nd and 50th centiles agree well, but the 98th centiles are more discordant. This is
probably because the z-score-conversion applies the UK90 skewness adjustment to the MWS data,
but the degree of ages. For BMI, the 2nd and 50th centiles agree well, but the 98th centiles are more discordant. This is
probably because the z-score conversion applies the UK90 skewness adjustment to the MWS data,
but the degree of

put the degree of skewness in the MWS data is different. With right skewness this would tend to
affect the upper rather than the lower centiles of the distribution, as seen here.
MWS centiles have already been published by

affect the upper rather than the lower centiles of the distribution, as seen here.

MWS centiles have already been published by the MWS Consortium(3), but superficially they lool

different from the centiles in Figure 4. T MWS centiles have already been published by the MWS Consortium(3), but supe
different from the centiles in Figure 4. This may be because we excluded differe
a different model-see Supplement 1 for our gamlss code.
There are different from the centiles in Figure 4. This may be because we excluded different outliers and fitt
a different model-see Supplement 1 for our gamlss code.
There are no published growth charts for ANKRD11, KMT2A, DDX3X or a different model-see Supplement 1 for our gamlss code.
There are no published growth charts for ANKRD11, KMT2A, DDX3X or ASXL3. Variants in ANKRD11
cause KBG Syndrome (MIM # 148050) (13). Our growth charts are consistent There are no published growth charts for *ANKRD11*, *KMT2*
cause KBG Syndrome (MIM # 148050) (13). Our growth ch
phenotypes. Growth hormone has been trialled as a thera
individuals with short stature, a review of which is There are no published growth charts for ANKRD11, KWT2A, DDX3X or ASXL3. Vandinis in ANKRD11
cause KBG Syndrome (MIM # 148050) (13). Our growth charts are consistent with the published
phenotypes. Growth hormone has been t phenotypes. Growth hormone has been trialled as a therapeutic option for a small number of
individuals with short stature, a review of which is underway by Dr Ockeloen's group (Aukema S.
al, submitted). These growth charts phenotypes. The individuals with short stature, a review of which is underway by Dr Ockeloen's group (Aukema
al, submitted). These growth charts will be an essential tool in monitoring children both in
determining need for individuals with short stature, a review of which is underway by Dr Duncom good, (and it is underway by Dr Ock
al, submitted). These growth charts will be an essential tool in monitoring children both in
determining need f determining need for potential treatment and judging its success.
determining need for potential treatment and judging its success. determining need for potential treatment and judging its success.

We selected DDXDX syndrome (MIM #300160), a incarded reportemental disorder predominantly in
females. Our charts agree with the literature of fairly normal heights and weights in females, with a
proportion having borderli females. Our charts agree with the literature of fairly normal heights and weights in females. Our charts agree with the literature of fairly normal heights and weights in female proportion having borderline microcephaly. es, with a
od and th
considere
e data po
WSS is
reight bel_i
g median proportion having borderline microcephaly. The phenotype in boys is less well understood and the
data are limited. Nonetheless our method enables some form of chart to be produced. Considered
with caution this could still

data are limited. Nonetheless our method enables some form of chart to be produced. Considered
with caution this could still provide a useful adjunct in a clinical setting, particularly if the data point
are viewable in th with caution this could still provide a useful adjunct in a clinical setting, particularly if the data point
are viewable in the chart to alert the clinician to the small sample size.
Alterations in *KMT2A* cause Wiedemann are viewable in the chart to alert the clinician to the small sample size.

Alterations in *KMT2A* cause Wiedemann-Steiner Syndrome (WSS) (MIM # 605130 (15). WSS is

associated with short stature in about 60% of individual Alterations in *KMT2A* cause Wiedemann-Steiner Syndrome (WSS) (MIN
associated with short stature in about 60% of individuals, microcephaly
the 5th centile in a third. Our charts are concordant with published desc
height

associated with short stature in about 60% of individuals, microcephaly in a third, and weight the 5th centile in a third. Our charts are concordant with published descriptions depicting medi
height, weight, and OFC for the 5th centile in a third. Our charts are concordant with published descriptions depicting median
height, weight, and OFC for *KMT2A* at the reference 2nd, 2nd and 25th centiles respectively.
ASXL3-related disorde the 5" centile in a third. Our charts are concordant with published descriptions depicting median
height, weight, and OFC for *KMT2A* at the reference 2nd, 2nd and 25th centiles respectively.
ASX13-related disorder (height, weight, and OFC for KMT2A at the reference 2nd, 2nd and 25th centiles respectively.
ASXL3-related disorder (Bainbridge-Ropers Syndrome, MIM # 615485) is associated with n
birth weight but poor postnatal growt Birth weight but poor postnatal growth due to feeding issues in infancy, which stabilise following
feeding intervention (16-18). BMI rises with age which may be explained by sustained feeding
intervention or because of dys beeding intervention (16-18). BMI rises with age which may be explained by sustained feeding
intervention or because of dysregulated or impulsive eating behaviours that can develop later in
childhood. Feeding interventions intervention or because of dysregulated or impulsive eating behaviours that can develop later
childhood. Feeding interventions should not target reference height (ASXL3 median height = 1
centile) and our charts will be imp

childhood. Feeding interventions should not target reference height (*ASXL3* median height = 12^{tt}
centile) and our charts will be important for growth monitoring. Median OFC is on the 3rd centile
consistent with publis childhood. Feeding interventions should not target reference height (ASXL3 median height = 12th
centile) and our charts will be important for growth monitoring. Median OFC is on the 3rd centile,
consistent with publish centile) and our charts will be important for growth monitoring. Median OFC is on the 3rd centile,
consistent with published literature of postnatal microcephaly.
Pre- and postnatal growth restriction, sometimes with mic Pre- and postnatal growth restriction, sometimes with microcephaly.

Individuals with variants in *SATB2* (19, 20) (MIM # 608148). Gro

(1508 data points) but they are hard to compare with ours as th

with chromosomal micr Individuals with variants in *SATB2* (19, 20) (MIM # 608148). Growth charts were produced recently
(1508 data points) but they are hard to compare with ours as they included single variants togethe
with chromosomal microde individuals with variants in SATB2 (19, 20) (MIM # 608140). Growth charts were produced recently
(1508 data points) but they are hard to compare with ours as they included single variants together
with chromosomal microdel (1508 data points) are hard to compute the total of y mustake inger taking together with chromosomal microdeletions; and they stopped at age 10 (5). They presented z-score heatmap
which indicate that the microdeletion grou which indicate that the microdeletion group was most affected by growth restriction, and this
probably explains the difference with our charts.
16 probably explains the difference with our charts. probably explains the difference with our charts.

Limitations

collated from various sources and some outliers may be due to measurement error, though we
sought out extreme measurements and either corrected or excluded them. Fewer OFC
measurements were available and birth OFC measurem sought out extreme measurements and either corrected or excluded them. Fewer OFC
measurements were available and birth OFC measurements were frequently missing. UK90 was
as the baseline reference for three reasons: most of sought out and will able and birth OFC measurements were frequently missing. U
as the baseline reference for three reasons: most of the data were from the UK (except
European ANKRD11 and MOWS and International ASXL3 cohort measurements were erasted and birth OFC measurements were from the UK (except for the
European ANKRD11 and MOWS and International ASXL3 cohorts), UK90 is the official UK growth
reference for data at birth and over 4 years European ANKRD11 and MOWS and International ASX*L3* cohorts), UK90 is the official UK growt
reference for data at birth and over 4 years of age (21), and the DECIPHER database uses UK90
display the growth data. However, th European AnkRD11 and MOWS and International ASALS conotts), onso is the official OK growth
reference for data at birth and over 4 years of age (21), and the DECIPHER database uses UK90 to
display the growth data. However,

reference for any characteristic for the state for data. However, there are inter-country differences in height for any child – see
Supplement 2 – and this is true for neurodevelopmental disorders as shown with ANKRD11 - c Supplement 2 – and this is true for neurodevelopmental disorders as shown with ANKRD11 - clinics
interpretation needs to take this into account.
The LMS2 method converts raw data to 2-scores using the baseline reference, Interpretation needs to take this into account.
The LMSz method converts raw data to 2-scores using the baseline reference, and then models the z
scores. It assumes that any skewness in the raw data matches the skewness i The LMSz method converts raw data to z-score
scores. It assumes that any skewness in the rav
this may not be the case. If the two skewness p
4B for BMI, where the 98th centiles based on th
particularly in males. BMI is m scores. It as may not be the case. If the two skewness patterns differ it will introduce bias, as seen in Figur
4B for BMI, where the 98th centiles based on the LMS and LMSz methods are very different,
particularly in ma

Implementation

alongside further growth measurements being collated through the GenROC study (7). Gene-specific This may note the 98th centiles based on the LMS and LMSz methods are very different,
particularly in males. BMI is markedly right skew at older ages (i.e. the L value is large and negative)
which affects the 98th cent 4B for BMI, where the 98" centiles based on the LMS and LMSz methods are very different,
particularly in males. BMI is markedly right skew at older ages (i.e. the L value is large and ne
which affects the 98th centile m particularly in males. Barricle in males in particular of the L value is large and negative),
which affects the 98th centile more than the $2nd$ and 50th centiles.
Implementation
We plan to automate LMSz within t which affects the 98" centile more than the 2"" and 50" centiles.
Implementation
We plan to automate LMSz within the DECIPHER platform (22, 23
alongside further growth measurements being collated through tl
growth charts w Mergen to antitude anti-mathematic plantinic (22, 23) and the GenROC study (7). Gene-sper
alongside further growth measurements being collated through the GenROC study (7). Gene-sper
growth charts within our proposed pipel along the further growth charts will be viewable within DECIPHER. A future development would be to variant subtype
specific charts within our proposed pipeline.
Conclusions
Children with genetic syndromes often have abnorm

Conclusions

specific charts within our proposed pipeline.
Conclusions
Children with genetic syndromes often have abnormal growth. LMSz allows clinically useful growth
charts to be generated from small datasets. These growth charts are Conclusions
Conclusions
Children with genetic syndromes often have a
charts to be generated from small datasets. T Charts to be generated from small datasets. These growth charts are gene specific which are
1 charts to be generated from small datasets. These growth charts are gene specific which are

Data availability

essential for accurate these charts to be available in the future for all paediatric
disorder genes.
Data availability
Sequence and variant-level data and phenotypic data for the DDD study data are available from the
Europ disorder genes.
Data availability
Sequence and variant-level data and phenotypic data for the DDD study data are available fro
European Genome-phenome Archive (EGA; https://www.ebi.ac.uk/ega/) with study ID
EGAS00001000775 Data availabili
Sequence and v:
European Genoi
EGAS000010007
and GenROC stu
Acknowledge [si H c A 、

Acknowledgements

European Genome-phenome Archive (EGA; https://www.ebi.ac.uk/ega/) with study ID
EGAS00001000775. Clinically interpreted variants and associated phenotypes from the DDD study
and GenROC study are available through DECIPHER EGA500001000775. Clinically interpreted variants and associated phenotypes from the
and GenROC study are available through DECIPHER (https://www.deciphergenomics.or
Acknowledgements
We thank all the individuals in the DDD and GenROC study are available through DECIPHER (https://www.deciphergenomics.org/).
Acknowledgements
We thank all the individuals in the DDD study, GenROC study and ASXL3 Natural History study and
other cohorts who consen and Centro-Centry are arbitrary through DDC study (https://) in through DECIR (https://).
Acknowledgements
We thank all the individuals in the DDD study, GenROC study and ASXL3 Natural History studther cohorts who consente We thank all the individuals in the DDD study, Gennoe study and ASXL3 Natural History study and
other cohorts who consented for anonymised data sharing. We thank Lucy Pollock (a member of
the GenROC study PPI group) for dr

Growth Chart Consortium for providing us with their dataset. We thank Lucy Pollock (a member of
the GenROC study PPI group) for drawing the original artwork in the graphical abstract.
The authors of this publication are me

The authors of this publication are members of the European Reference Network on Rare Congenition
The authors of this publication are members of the European Reference Network on Rare Congenition
Malformations and Rare Int The authors of this publication are members of the European Reference Network on Ram

Malformations and Rare Intellectual Disability ERN-ITHACA [EU Framework Partnership

ID: 3HP-HP-FPA ERN-01-2016/739516].

The GenROC stu Malformations and Rare Intellectual Disability ERN-ITHACA [EU Framework Partnership Agreement
ID: 3HP-HP-FPA ERN-01-2016/739516].
The GenROC study would not be possible without the GenROC consortium which is comprised
mult Malformations and Malformations and Rare Intellectual Disability Partnership Agreement
The GenROC study would not be possible without the GenROC consortium which is comprised
multiple contributors from NHS sites across the The GenROC study would not be possibl
multiple contributors from NHS sites aci
have either acted as a local site PI or As.
list of individuals is as follows: Suzanne.
Diana Baralle, Jonathen Berg , Marta Be
Helen Brittain, The Contributors from NHS sites across the UK. The consortium is comprised of individuals
have either acted as a local site PI or Associate PI or completed clinical proformas for GenROC.
list of individuals is as follows: have either acted as a local site PI or Associate PI or completed clinical proformas for GenROC. The
list of individuals is as follows: Suzanne Alsters, Ruth Armstrong, Tazeen Ashraf, Queenstone Baker,
Diana Baralle, Jonat ist of individuals is as follows: Suzanne Alsters, Ruth Armstrong, Tazeen Ashraf, Queenstone Baker,
Diana Baralle, Jonathen Berg , Marta Bertoli, Thomas Boddington, Moira Blyth, Catherine Breen,
Helen Brittain, Lisa Bryson Diana Baralle, Jonathen Berg, Marta Bertoli, Thomas Boddington, Moira Blyth, Catherine Breen,
Helen Brittain, Lisa Bryson, Jenny Carmichael, Emma Clement, Tessa Coupar, Anna de Burca, Cristina
Dias, Fleur Dijk, Abhijit Dix Helen Brittain, Lisa Bryson, Jenny Carmichael, Emma Clement, Tessa Coupar, Anna de Burca, Crist
Dias, Fleur Dijk, Abhijit Dixit, Alan Donaldson, Andrew Douglas, Jacqueline Eason, Fayadh Fauzi,
Dias, Fleur Dijk, Abhijit Dix Dias, Fleur Dijk, Abhijit Dixit, Alan Donaldson, Andrew Douglas, Jacqueline Eason, Fayadh Fauzi,
18
18 Dias, Fleur Dijk, Abhijit Dixit, Alan Donaldson, Andrew Douglas, Jacqueline Eason, Fayadh Fauzi,

Gardner, Merrie Gowie, Rachel Harrison, Verity Hartill, Lizzie Harris, Eleanor Hay, Jenny Higgs, Simo
Holden, Daniela Iancu, Rachel Irving, Vani Jain, Rosalyn Jewell, Gabriela Jones, Beckie Kaemba,
Arveen Kamath, Ayse Nur Holden, Daniela lancu, Rachel Irving, Vani Jain, Rosalyn Jewell, Gabriela Jones, Beckie Kaemba,
Arveen Kamath, Ayse Nur Kavasoglu, Mira Kharbanda, Sophie King, Alison Kraus, Ajith Kumar,
Katherine Lachlan, Neeta Lakhani, W Arveen Kamath, Ayse Nur Kavasoglu, Mira Kharbanda, Sophie King, Alison Kraus, Ajith Kumar,
Katherine Lachlan, Neeta Lakhani, Wayne Lam, Anne Lampe, Abigail Lazenbury, Helen Leveret,
Jessica Maiden, Alison Male, Alisdair Mc Aratherine Lachlan, Neeta Lakhani, Wayne Lam, Anne Lampe, Abigail Lazenbury, Helen Leveret, Jessica Maiden, Alison Male, Alisdair McNeil. Ruth McGowan, Holly McHale, Catherine McWill
Jonathan Memish, Lara Menzies, Radwa Mo Resica Maiden, Alison Male, Alisdair McNeil. Ruth McGowan, Holly McHale, Catherine McWilli
Jonathan Memish, Lara Menzies, Radwa Mohamed, Tara Montgomery, Oliver Murch, Michael
Parker, Caroline Pottinger, Vijayalakshmi Rama Jonathan Memish, Lara Menzies, Radwa Mohamed, Tara Montgomery, Oliver Murch, Michael
Parker, Caroline Pottinger, Vijayalakshmi Ramakumaran, Ruth Richardson, Alison Ross, Claire Searle
Charles Shaw-Smith, Suresh Somarathi, Parker, Caroline Pottinger, Vijayalakshmi Ramakumaran, Ruth Richardson, Alison Ross, Claire S
Charles Shaw-Smith, Suresh Somarathi, Edward Steel, Helen Stewart, Kerra Templeton, Riya
Tharakan, Madeline Tooley, Mohamed Wafi ranch, caroline Pottinger, Vijayalakshmi Ramakumaran, Ruth Richardson, Alison Ross, Claire Searle,
Charles Shaw-Smith, Suresh Somarathi, Edward Steel, Helen Stewart, Kerra Templeton, Riya
Tharakan, Madeline Tooley, Mohamed

Funding statement

Tharakan, Madeline Tooley, Mohamed Wafik, Emma Wakeling, Elizabeth Wall, Amy Watford
Patricia Wells, Louise Wilson
Funding statement
The DDD study presents independent research commissioned by the Health Innovation Chall
F Patricia Wells, Louise Wilson
Tharakan, Madeline Tooley, Mohamed Waffel, Innovation Challe
The DDD study presents independent research commissioned by the Health Innovation Challe
The Department of Health, and the Wellcome Funding statement
The DDD study presents inde
Fund [grant number HICF-100
the Department of Health, an
was supported by the Nation:
Centre. The research team ac | | | F | t | C | t | r Fund [grant number HICF-1009-003], a parallel funding partnership between the Wellcome Trust at
the Department of Health, and the Wellcome Sanger Institute [grant number WT098051]. This stu
was supported by the National In Fund Department of Health, and the Wellcome Sanger Institute [grant number WT098051]. This study
was supported by the National Institute for Health and Care Research Exeter Biomedical Research
Centre. The research team ack the Department of Netalth and Care Research Exeter Biomedical Research
Centre. The research team acknowledges the support of the National Institute for Health Research,
through the Comprehensive Clinical Research Network. Centre. The research team acknowledges the support of the National Institute for Health Research
through the Comprehensive Clinical Research Network. JF is funded by the Wellcome Trust [grant
number WT223718/Z/21/Z]. This through the Comprehensive Clinical Research Network. JF is funded by the Wellcome Trust [grant
number WT223718/Z/21/Z]. This research was funded in whole or in part by the Wellcome Trust.
MB is funded by the MRC (MR/V03730 number WT223718/Z/21/Z]. This research was funded in whole or in part by the Wellcome Trust.
NB is funded by the MRC (MR/V037307/1). For the purpose of open access, the author has applied
CC-BY public copyright licence to MB is funded by the MRC (MR/V037307/1). For the purpose of open access, the author has applie
CC-BY public copyright licence to any author accepted manuscript version arising from this
submission. KL is supported by the Na CC-BY public copyright licence to any author accepted manuscript version arising from this
submission. KL is supported by the National Institute for Health and Care Research Doctoral
Research Fellowship 302303. The views e Examples in KL is supported by the National Institute for Health and Care Research Doctora
Research Fellowship 302303. The views expressed are those of the author(s) and not neces
those of the NIHR or the Department of Hea Research Fellowship 302303. The views expressed are those of the author(s) and not necess
those of the NIHR or the Department of Health and Social Care.
Those of the NIHR or the Department of Health and Social Care. Research Fellowship 3023303. The views expressed are those of the author(s) and not necessary)
those of the NIHR or the Department of Health and Social Care. those of the NIHR or the Department of Health and Social Care.

ノ く ド ト ト Author contributions

Ethics declaration

Formal analysis: TJC
Project administration, Supervision, Writing-original draft: KJL
Data curation, Writing-review & editing: all authors.
Ethics declaration
The DDD study has UK Research Ethics Committee approval (: Project administration
Data curation, Writir
Ethics declaration
The DDD study has L
Cambridge South RE
informed consent, as Provide a Constant and Solar Ethics declaration
Proposition, The DDD study has UK Research Ethics Committee approval (10/HO
Cambridge South REC, and GEN/284/12 gr Ethics declaration
The DDD study has UK Research Ethics Committee ap
Cambridge South REC, and GEN/284/12 granted by t
informed consent, as required by the REC. All publish
approval for the growth charts development was giv トー くしこん くし Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC). All participation
informed consent, as required by the REC. All published data were de-identified. Specific et
approval for the growth charts dev Entering Consent, as required by the REC. All published data were de-identified. Specific ethical
approval for the growth charts development was given via a DDD Complementary Analysis Proposal
Approval (CAP#371). The GenRO informed consent for the growth charts development was given via a DDD Complementary Analysis Prope
Approval (CAP#371). The GenROC study received East Midlands - Nottingham Research Ethics
Committee (REC) approval on 15 De Approval (CAP#371). The GenROC study received East Midlands - Nottingham Research Ethics
Committee (REC) approval on 15 December 2022 and Health Research Authority approval on 9
February 2023. The ASXL3 Natural History Stu Committee (REC) approval on 15 December 2022 and Health Research Authority approval on 9
February 2023. The ASXL3 Natural History Study, sponsored by Sheffield Children's Hospital an
University of Sheffield (UK) received R Committee (REC) approval on 15 December 2022 and Authority approval on 2 June 2023. The ASXL3 Natural History Study, sponsored by Sheffield Children's Hospital and University of Sheffield (UK) received REC (23/SC/0151) and February 2023. The ASXL3 Natural History Study, sponsored by Sheffield Children's Hospital and The
University of Sheffield (UK) received REC (23/SC/0151) and HRA approval on 2 June 2023. All
participants enrolled in the st participants enrolled in the study gave informed consent for anonymised data sharing to allo
collaboration.
Conflict of interest statement
None declared. participants enrolled in the study gave information.

Conflict of interest statement

None declared.

Conflict of ir
None declared
None declared $\frac{1}{2}$ Conflict of interest statement

References:

 $1₁$

monitoring as an early detection tool: a systematic review. The Lancet Diabetes & Endocrinol
2016;4(5):447-56.
2. Shur N, Tigranyan A, Daymont C, Regier DS, Raturi S, Roshan Lal T, et al. The past, pres
and future of child 2016;4(5):447-56.

2. Shur N, Tigranyan A, Daymont C, Regier DS, Raturi S, Roshan Lal T, et al. The past, present,

2. Shur N, Tigranyan A, Daymont C, Regier DS, Raturi S, Roshan Lal T, et al. The past, present,

2023;191(2. Shur N, Tig
and future of child
2023;191(4):948-6
3. Ivanovski I
syndrome: growth
4. McCague I
Coffin-Siris syndron
5. Zarate YA,
syndrome. Am J M
6. Firth HV, V
Developmental Me
7. Low KJ, Wi
with GENetic Rare

and future of child growth monitoring: A review and primer for clinical genetics. Am J Med Genet A
2023;191(4):948-61.
2023;191(4):948-61.
3. Ivanovski I, Djuric O, Broccoli S, Caraffi SG, Accorsi P, Adam MP, et al. Mowat-2023;191(4):948-61.

3. Ivanovski I, Djuric O, Broccoli S, Caraffi SG, Accorsi P, Adam MP, et al. Mowat-Wilson

syndrome: growth charts. Orphanet J Rare Dis. 2020;15(1):151.

4. McCague EA, Lamichhane R, Holt N, Schrier Ve 3. Ivanovski I, D
Syndrome: growth ch
4. McCague EA,
Coffin-Siris syndrome
5. Zarate YA, Ka
syndrome. Am J Med
6. Firth HV, Wri
Developmental Medi
7. Low KJ, Watf
with GENetic Rare di:
2024;14(5):e085237.
8. Cole TJ, Gree
 syndrome: growth charts. Orphanet J Rare Dis. 2020;15(1):151.

4. McCague EA, Lamichhane R, Holt N, Schrier Vergano SA. Growth charts for individual

Coffin-Siris syndrome. Am J Med Genet A. 2020;182(10):2253-62.

5. Zarat Syndemer.growth charts or prothers into a strategies (applied SCOffin-Siris syndrome. Am J Med Genet A. 2020;182(10):2253-6;

23 Zarate YA, Kannan A, Bosanko KA, Caffrey AR. Growth in

syndrome. Am J Med Genet A. 2022;188(Coffin-Siris syndrome. Am J Med Genet A. 2020;182(10):2253-62.

2. Zarate YA, Kannan A, Bosanko KA, Caffrey AR. Growth in individuals with SATB2-associated

syndrome. Am J Med Genet A. 2022;188(10):2952-7.

6. Firth HV, Wr Example Protection and Alexander Material Care (Associated VA, Care (And Syndrome. Am J Med Genet A. 2022;188(10):2952-7.

6. Firth HV, Wright CF, DDD Study. The Deciphering Develop

Developmental Medicine & Child Neurolog 5. Firth HV, Wright CF, DDD Study. The Deciphering Developmental Disorders (DDD) study.

5. Firth HV, Wright CF, DDD Study. The Deciphering Developmental Disorders (DDD) study.

5. Low KJ, Watford A, Blair P, Nabney I, Pow Firth HV, Wright CF, DDD Study. The Decipher
Developmental Medicine & Child Neurology. 2011;53
7. Low KJ, Watford A, Blair P, Nabney I, Powell J
with GENetic Rare disease: Observational Cohort stuc
2024;14(5):e085237.
8. C Developmental Medicine & Child Neurology. 2011;53(8):702-3.

7. Low KJ, Watford A, Blair P, Nabney I, Powell J, Wynn SL, et al. Improving the care of child

with GENetic Rare disease: Observational Cohort study (GenROC)—a T. Low KJ, Watford A, Blair P, Nabney I, Powell J, Wynn SL,
with GENetic Rare disease: Observational Cohort study (GenROO
2024;14(5):e085237.
8. Cole TJ, Green PJ. Smoothing reference centile curves: t
likelihood. Stat Med

with GENetic Rare disease: Observational Cohort study (GenROC)—a study protocol. BMJ Open.
2024;14(5):e085237.
8. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized
likelihood. Stat Med. 19 2024;14(5):e085237.

8. Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized

1. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape.

1. Rigby RA, Stasinopo 2022;16(2):2022;16(2):2022;16(2):2022;16(2):2022;16(2):2022;116(1):2022;116(1):122-31.
2022;116(1):2022;116(1):2022;116(1):2022;116(1):122-31.
2022;116(1):122-31.
2022;116(1):122-31.
13. Low K, Ashraf genetic aspects of KB ikelihood. Stat Med. 1992;11(10):1305-19.
9. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape
1000rmal of the Royal Statistical Society: Series C (Applied Statistics). 2005;54(3):507-54 9. Rigby RA, Stasinopoulos DM. Gener
Journal of the Royal Statistical Society: Serie
10. Cole TJ. Sample size and sample cor
Statistical Methods in Medical Research. 20
11. Cole TJ, Freeman JV, Preece MA. Bo
Dis Child. 199 1020 Journal of the Royal Statistical Society: Series C (Applied Statistics). 2005;54(3):507-54.

10. Cole TJ. Sample size and sample composition for constructing growth reference centiles.

5. Statistical Methods in Medic Statistical Methods in Medical Research. 2021;30(2):488-507.

11. Cole TJ, Freeman JV, Preece MA. Body mass index reference curves for the UK, 1990. Arc

Dis Child. 1995;73(1):25-9.

12. Wright CM, Cole TJ, Fewtrell M, Wil

Cole TJ. Sample size and sample composition for constructing growth reference
Statistical Methods in Medical Research. 2021;30(2):488-507.
11. Cole TJ, Freeman JV, Preece MA. Body mass index reference curves for the UK,
Di Statistical Methods in Methods in Methods in Methods in Methods in Solidian 1995;73(1):25-9.
12. Mright CM, Cole TJ, Fewtrell M, Williams JE, Eaton S, V
that high BMI centiles overdiagnose obesity in children aged u
2022;1 12. Wright CM, Cole TJ
that high BMI centiles over
2022;116(1):122-31.
13. Low K, Ashraf T, Ca
genetic aspects of KBG synd
14. Levy T, Siper PM, LG
Summary of Findings and R
2023;138:87-94.
15. Jones WD, Dafou D
Novo Mutat 2022;116(1):122-31.

13. Low K, Ashraf T, Canham N, Clayton-Smith J, Deshpande C, Donaldson A, et al. Clini

14. Levy T, Siper PM, Lerman B, Halpern D, Zweifach J, Belani P, et al. DDX3X Syndrome

Summary of Findings and R 2022;116(1):122-31. examples and Recommentations of Evaluation and Care. Pediatric Levy T, Siper PM, Lerman B, Halpern D, Zweifach J, Belani P, et al. DDX
Summary of Findings and Recommendations for Evaluation and Care. Pediatri
2023;138:87-9

11. Cole TJ, Freeman Maria Maria Maria Maria Maria Maria Maria Maria Maria Shotical and that high BMI centiles overdiagnose obesity in children aged under 6 years. Am J Clin Nutr.

12. Wright CM, Cole TJ, Fewtrell M, Willi that high BMI centiles overdiagnose obesity in children aged under 6 years. Am J Clin Nutr.
2022;116(1):122-31.
13. Low K, Ashraf T, Canham N, Clayton-Smith J, Deshpande C, Donaldson A, et al. Clinical and
genetic aspects Experie aspects of KBG syndrome. Am J Med Genet A. 2016;170(11):2835-46.

14. Levy T, Siper PM, Lerman B, Halpern D, Zweifach J, Belani P, et al. DDX3X Syndrome:

Summary of Findings and Recommendations for Evaluation and Summary of Findings and Recommendations for Evaluation and Care. Pediatric Neurology.
2023;138:87-94.
2023;138:87-94.
15. Jones WD, Dafou D, McEntagart M, Woollard WJ, Elmslie FV, Holder-Espinasse M, et
Novo Mutations in M 2023;138:87-94.

Summary of Findings and Nullel Multimary of Findings EV, Holder-Espinasse M, 6

Novo Mutations in MLL Cause Wiedemann-Steiner Syndrome. The American Journal of Hur

Senetics. 2012;91(2):358-64.

Bainbridge 15. Jones WI
Novo Mutations
Genetics. 2012;9.
16. Bainbridg
truncating mutat
Bohring-Opitz syr

Novo Mutations in MLL Cause Wiedemann-Steiner Syndrome. The American Journal of Human
Genetics. 2012;91(2):358-64.
16. Bainbridge MN, Hu H, Muzny DM, Musante L, Lupski JR, Graham BH, et al. De novo
truncating mutations in Genetics. 2012;91(2):358-64.
16. Bainbridge MN, Hu H, Muzny DM, Musante L, Lupski JR, Graham BH, et al. De novo
truncating mutations in ASXL3 are associated with a novel clinical phenotype with similarities to
Bohring-Opit The Bainbridge MN, Hu H,
truncating mutations in ASXL3
Bohring-Opitz syndrome. Gen 16. Bainbridge May, Hunter, John Museum, Hungarian, John Museum, John Museum Litruncating mutations in ASXL3 are associated with a novel clinical phenotype with similaritie
Bohring-Opitz syndrome. Genome Med. 2013;5(2):11. Bohring Opitz syndrome. Genome Med. 2013;5(2):11. Bohring-Opitz syndrome. Genome Med. 2013;5(2):11.

17.

Delineating the phenotypic spectrum of Bainbridge-Ropers syndrome: 12 new patients with
terozygous, loss-of-function mutations in ASXL3 and review of published literature. J M:
2017;54(8):537-43.
18. Schirwani S, Albaba S, heterozygous, loss-of-function mutations in ASXL3 and review of published literature. J Med Genet.

2017;54(8):537-43.

18. Schirwani S, Albaba S, Carere DA, Guillen Sacoto MJ, Milan Zamora F, Si Y, et al. Expanding

the p 2017;54(8):537-43.

18. Schirwani S, Albaba S, Carere DA, Guillen Sacoto MJ, Milan Zamora F, Si Y, et al. Expanding

the phenotype of ASXL3-related syndrome: A comprehensive description of 45 unpublished

individuals with 18. Schirwani S,
the phenotype of A!
individuals with inhe
2021;185(11):3446-
19. Bengani H, I
consequences of dis
2017;19(8):900-8.
20. Döcker D, S!
the SATB2 phenoty
21. Wright CM,
WHO growth charts
22. Foreman J,
Suppor the phenotype of ASXL3-related syndrome: A comprehensive description of 45 unpublished
individuals with inherited and de novo pathogenic variants in ASXL3. Am J Med Genet A.
2021;185(11):3446-58.
19. Bengani H, Handley M,

individuals with inherited and de novo pathogenic variants in ASXL3. Am J Med Genet A.
2021;185(11):3446-58.
19. Bengani H, Handley M, Alvi M, Ibitoye R, Lees M, Lynch SA, et al. Clinical and molecu
consequences of disease 2021;185(11):3446-58.

19. Bengani H, Handley M, Alvi M, Ibitoye R, Lees M, Lynch SA, et al. Clinical and mol

2021;185(11):3446-58.

19. Bengani H, Handley M, Alvi M, Ibitoye R, Lees M, Lynch SA, et al. Clinical and mol

2022, Bengani H, Han
2017;19(8):900-8.
2017;19(8):900-8.
20. Döcker D, Schulthe SATB2 phenotype. E
21. Wright CM, Will
WHO growth charts. BN
22. Foreman J, Brens Supporting the interprediagnosis and research.
23. Firth HV, 19. Bengani H, Bengani
19. Bengani H, Saray D, Schubach M, Menzel M, Munz M, Spaich C, Biskup S, et al. Further delineatio
19. Bendi H 2017;19(8):900-8.

2017;19(8):900-8.

20. Döcker D, Schubach M, Menzel M, Munz M, Spaich C, Biskup S, et al. Further del

21. Wright CM, Williams AF, Elliman D, Bedford H, Birks E, Butler G, et al. Using the 1

21. Wright 2019, 2019, 2019
20. Döcker D,
the SATB2 phenoty
21. Wright CM
WHO growth chart
22. Foreman J
Supporting the interaliagnosis and rese
23. Firth HV, R
Chromosomal Imb
human genetics. 2
Legends 20. Unit SATB2 phenotype. European Journal of Human Genetics. 2014;22(8):1034-9.

21. Wright CM, Williams AF, Elliman D, Bedford H, Birks E, Butler G, et al. Using the new UK-WHO growth charts. BMJ. 2010;340:c1140.

22. Fo 21. Wright CM, Williams AF, Elliman D, Bedford H, Birks E, Butler G, et al. Usit WHO growth charts. BMJ. 2010;340:c1140.
22. Foreman J, Brent S, Perrett D, Bevan AP, Hunt SE, Cunningham F, et al. DESupporting the interpre Example interpretation and sharing of
Supporting the interpretation and sharing of
diagnosis and research. Human mutation. 21
23. Firth HV, Richards SM, Bevan AP, Clachronosomal Imbalance and Phenotype in
human genetics. 2 Supporting the interpretation and sharing of rare disease phenotype-linked variant data to
diagnosis and research. Human mutation. 2022;43(6):682-97.
23. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al diagnosis and research. Human mutation. 2022;43(6):682-97.
23. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of
Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resourc 23. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M,
Chromosomal Imbalance and Phenotype in Humans Using Ens
human genetics. 2009;84 4:524-33.
Legends
Figure 1: mean ±2 SD lines representing the 2nd, 50th and 98^t 23. Figure 1: mean ±2 SD lines representing the 2nd, 50th and 98th gene-specific z-score "centiles" for height, weight, BMI and OFC across the genes *ANKRD11, ARID1B, ASXL3, DDX3X, KMT2A and SATB2*.
The corresponding human genetics. 2009;84 4:524-33.

Legends

Figure 1: mean ±2 SD lines representing the 2nd, 50th and 98th gene-specific z-score "centiles" for

height, weight, BMI and OFC across the genes *ANKRD11, ARID1B, ASXL3, D*

| F F F F Legends

human generation 2009;

Legends

Figure 1: mean ±2 SD lines represen

height, weight, BMI and OFC across

The corresponding UK90 z-score cer

Figure 2: measurement centiles for t

transforming the z-score centiles sh

height, weight, BMI and OFC across the genes ANKRD11, ARID1B, ASXL3, DDX3X, KMT2A and SATB2.

Figure 2: measurement centiles for females by measurement and gene obtained b
transforming the z-score centiles shown in Figure 1.
Figure 3: measurement centiles for males by measurement and gene obtained by
the z-score c

Figure 1: mean ±2 SD lines representing the 2nd, 50th and 98th gene-specific z-score "centiles" for
height, weight, BMI and OFC across the genes *ANKRD11, ARID1B, ASXL3, DDX3X, KMT2A and SA*
The corresponding UK90 zheight, weight, BMI and OFC across the genes ANKRD11, ARD11, ASXL3, DDX3X, KMT2A and SATD21.
The corresponding UK90 2-score centiles at -2, 0 and 2 are shown as dashed lines.
Figure 2: measurement centiles for females by m Figure 3: measurement centiles for males by measur
the z-score centiles shown in Figure 1.
Figure 4: A) MWS data on the z-score scale, in the sa
50th and 98th centiles by the two methods, for the se
LMS method and das

Figure 4: A) MWS data on the z-score s
50th and 98th centiles by the two metho
LMS method and dashed curves to the

Figure 4: A) MWS data on the z-score scale, in the same form as Figure 1. B) Comparison of the 2^{nd} ,
Figure 4: A) MWS data on the z-score scale, in the same form as Figure 1. B) Comparison of the 2^{nd} ,
50th and 98^t Figure 4: A) MWS data on the z-score scale, in the same form as Figure 1. B) Comparison of the 2nd,
50th and 98th centiles by the two methods, for the sexes separately - solid curves correspond to the
LMS method and 50th and 98th centiles by the two methods, for the sexes separately - solid curves correspond to the

Supplementary Figure 1: Z score "centiles" for ANKRD11 for OK data only.
Supplementary Figure 2: 4x6 plots as in Figure 1 but penalised with the AIC
Supplementary Figure 3: 4x6 plots as in Figure 1 but with the option of a

Supplementary Figure 3: 4x6 plots as in Figure 1 but with the option of a P-s
Supplement 1: MWS LMS method
Supplement2: Median z score for height for ANKRD11 by country
Supplementary Table 1 : regression coefficients and s Supplement 1: MWS LMS method
Supplementary Table 1 : regression coefficients and standard errors for the BIC models underlying
Figure 1; n = 53.
Supplementary Table 2 : regression coefficients and standard errors for the A

Supplement 2: Median z score for h
Supplementary Table 1 : regression
Figure 1; n = 53.
Supplementary Table 2 : regression
Supplementary Figure 2; n=74.
Supplementary Table 3: 2nd, 50th and The 2nd centile is report Supplementary Table 1 : regression coefficients and standard errors
Figure 1; n = 53.
Supplementary Table 2 : regression coefficients and standard errors
Supplementary Figure 2; n=74.
Supplementary Table 3: 2^{nd} , 50th Figure 1; n = 53.
Supplementary Table 2 : regression coefficients and standard errors for the AIC models underlying
Supplementary Figure 2; n=74.
Supplementary Table 3: 2^{nd} , 50th and 98th centiles at birth for heig Supplementary 1
Supplementary F
Supplementary 1
The 2^{nd} centile is Supplementary Figure 2; n=74.

Supplementary Table 3: 2^{nd} , 50th and 98th centiles at birth for height, weight, BMI and OFC by gene.

The 2nd centile is reported as a z-score for reasons explained. Supplementary Table 3: 2^{nd} , 50°
The 2^{nd} centile is reported as a : Supplementary Table 3: 2^{nc} , 50" and 98" centiles at birth for height, weight, BMI and OFC by gene.
The 2^{nd} centile is reported as a z-score for reasons explained. The 2nd centile is reported as a z-score for reasons explained.

centile $-$ 2nd $-$ 50th $-$ 98th

Female

centile $-$ 2nd $-$ 50th $-$ 98th

Male

A

