Fit to WHO weight standard of European infants over time

Daniel Levin,1 Louise Marryat, 1 Tim J Cole,2 John McColl,1 Ulla Harjunmaa,3,4 Per Ashorn,3,4 Charlotte Wright5

ABSTRACT
Objectives The 2006 WHO growth charts were created to provide an international standard for optimal growth, based on healthy, breastfed populations, but it has been suggested that Northern European children fit them poorly. This study uses infant weight data spanning 50 years to determine how well-nourished preschool children from different eras fit the WHO standard, and discuss the implications of deviations.

Design Four longitudinal datasets from the UK and one from Finland were used comprising over 8000 children born between 1959 and 2003. Weights from birth to 2 years were converted to age–sex-adjusted Z scores using the WHO standard and summarised using Generalized Additive Models for Location, Scale and Shape.

Results Weights showed a variable fit to the WHO standard. Mean weights for all cohorts were above the WHO median at birth, but dipped by up to 0.5 SD to a nadir at 8 weeks before rising again. Birth weights increased in successive cohorts and the initial dip became slightly shallower. By age 1 year, cohorts were up to 0.75 SD above the WHO median, but there was no consistent pattern by era.

Conclusions The WHO standard shows an acceptable, but variable fit for Northern European infants. While birth weights increased over time, there was, unexpectedly, no consistent variation by cohort beyond this initial period. Discrepancies in weight from the standard may reflect differences in measurement protocol and trends in infant feeding practice.

INTRODUCTION
Growth charts are widely used in child health to identify undernutrition and overnutrition.1 Many countries produce their own charts describing how local children grow. However, these do not necessarily characterise healthy growth; in some countries average weight will be low due to undernutrition, while in high-income countries higher average weight will be caused by rising rates of obesity. In the past, charts were based predominantly on bottle-fed infants and did not adequately reflect the growth of breastfed infants who should be the physiological norm.2 Thus, the WHO developed new growth charts based on healthy breastfed children living in optimal circumstances in six world regions.3 The WHO showed how all children aged 0–5 years should grow, whatever their ethnic origin. However, weights by country for the WHO dataset have never been published. Since the publication of the standard, studies in unselected, healthy, non-deprived populations such as the UK,7 USA,6 Canada,3,9 Norway, Belgium,8 Italy, Argentina9 and Denmark10 have generally found a close fit for length, but a tendency for children to become heavier than the WHO standard after the first 6 months. Some authors have argued that these discrepancies represent a fundamental difference that renders the standard unsuitable for high-income countries.9 Others have suggested that variations in fit are to be expected if the WHO charts represent optimal rather than average growth, as few children will be breastfed to age 1 year and that the higher weights reflect rising rates of obesity at all ages.6

If the mismatch is due to obesity, it ought to be less evident in historic datasets from eras when obesity was less common. Conversely, if it reflects the nature, prevalence and duration of formula feeding, the fit should be better in more recent cohorts as breastfeeding has increased.

Aims
Our aims are to explore a portfolio of weight datasets from the past 50 years (i) to determine how well real populations of well-nourished preschool children from different eras fit the 2006 WHO growth standard at different ages during infancy;
and (ii) to explore how trends differ by era of the cohort and associated infant feeding patterns.

**METHODS**

**Datasets**

Data came from existing longitudinal growth studies, retrieved mainly from routine records. They had already been cleaned, checked and analysed for other purposes, with four studies already published. Details of the five studies are as follows:

**Widdowson study (1959)**

In a study set up by Dr Elsie Widdowson, routine weights of 1094 babies born in 1959–1965 were obtained contemporaneously from the records of 10 Cambridge Child Welfare Clinics. Weights were measured by clinical staff approximately monthly in the first year of life, with a maximum of 13 weights per child. Of the possible 14 222 measurements all but 864 (6%) were collected. Although the data were cleaned and analysed at the time of collection the results were never published.

**Cambridge Infant Growth study (1984)**

Cambridge Infant Growth Study (CIGS) was a research cohort of 255 babies recruited in 1984–1987. Infants were sampled in four cohorts from lists of Cambridge city mothers booked to deliver in particular months, with some filtering by midwives. Measurements were taken mainly by one highly trained auxologist every 4 weeks from 4 to 52 weeks, then at 15, 18, 24, 30, 36 and 48 months. Weight, length, head and arm circumferences, triceps and subscapular skinfolds were measured at each visit: 223 (87%) had all 15 measurements from birth to 2 years.12–14

**Newcastle Growth and Development study (1987)**

This dataset comprises the routine weights of 2809 children aged 0–4 years born between October 2003 and September 2004 who attended child health clinics in Tampere. Children were weighed by clinical staff on electronic scales. Up to 16 scheduled events were recorded per child at birth, 1–2 weeks, 6–8 weeks, 2, 3, 4, 5, 6, 8, 10, 12, 18, 24, 36 and 48 months. There was a mean of 12 measurements per child.20

**Analysis**

The datasets were cleaned and weights converted into Z scores relative to the WHO growth standard. Data beyond 2 years were excluded, when numbers were low and bias was likely. The measurements at birth and 6–8 weeks, 3, 6, 9, 12 18 and 24 months were summarised by age, sex and cohort (table 1). The mean Z score, SD, skewness and kurtosis were modelled as functions of age for each dataset by sex, using Generalized Additive Models for Location, Scale and Shape (GAMLSS) as implemented in the GAMLSS package in R V.3.1.1. Multiple GAMLSS models were fitted using different hyperparameters and distribution families. In the final models, based on the Box-Cox power exponential family (BCPE), the mean Z score was allowed to vary with age in each of the datasets, whereas the SD, skewness and kurtosis were constrained to be constant as exploratory analyses showed this made little difference to the model fit as determined using the Bayesian Information Criterion. The skewness adjustment ensured that the mean and median were effectively the same. The BCPE requires values to be positive, so all Z scores had 10 added to them prior to analysis, and 10 was then subtracted from the mean curves. The

<table>
<thead>
<tr>
<th>Target age (weeks)</th>
<th>Age range</th>
<th>Sex</th>
<th>Number of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Birth</td>
<td>M</td>
<td>573</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>519</td>
</tr>
<tr>
<td>7</td>
<td>5–9 weeks</td>
<td>M</td>
<td>452</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>410</td>
</tr>
<tr>
<td>13</td>
<td>2–4 months</td>
<td>M</td>
<td>573</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>520</td>
</tr>
<tr>
<td>26</td>
<td>5–7 months</td>
<td>M</td>
<td>570</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>516</td>
</tr>
<tr>
<td>39</td>
<td>8–10 months</td>
<td>M</td>
<td>569</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>515</td>
</tr>
<tr>
<td>52</td>
<td>11–14 months</td>
<td>M</td>
<td>501</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>476</td>
</tr>
<tr>
<td>78</td>
<td>15–21 months</td>
<td>M</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>–</td>
</tr>
<tr>
<td>104</td>
<td>22–30 months</td>
<td>M</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>–</td>
</tr>
</tbody>
</table>

CIGS, Cambridge Infant Growth Study; GDS, Growth and Development Study; GMS, Gateshead Millennium Study.
models for each cohort were plotted as mean Z score versus age in boys and girls, along with the constant SD, skewness and kurtosis, to compare how well the cohorts fitted the WHO 2006 standard. A mean Z score of 0, SD 1, skewness 1 and kurtosis 2 indicates a perfect fit, whereas a mean Z score above 0 is heavier than the WHO standard, and below is lighter.

There is no standard definition of what constitutes a good or poor fit to a growth reference. In this study fit was defined in terms of mean weight Z score, measured in fractions of a centile channel width relative to the WHO median, where one channel width = 0.67 SDs. An excellent fit was defined as a difference of no more than ¼ of a channel width (0.17 SD) and a poor fit as greater than a channel width (0.67 SD).

RESULTS
Mean birthweight Z scores in the five datasets were all positive and close to zero (table 2). They were progressively higher in later years, particularly in the boys, rising from 0.03 in 1959 to 0.37 in 2003. For all cohorts the fit was adequate in early infancy, mostly staying within half a channel width of the median, but by 1 year boys and girls in Widdowson and boys in Tampere fitted poorly, more than a channel width above the median.

The smoothed curves of mean Z score versus age from the five datasets were plotted on separate charts by sex, along with the constant SD (sigma), skewness (nu) and kurtosis (tau) (figure 1). The curves provide a visual assessment of the fit to the standard by cohort. All datasets showed consistent differences relative to the standard. The mean Z scores all tended to fall in the early weeks, with the four UK datasets ending up below zero. This steep fall was followed by a slightly less steep rise, creating a ‘trough’ in each curve at around 8 weeks. Boys and girls followed broadly the same pattern.

The depth of the trough, like birthweight, became progressively shallower over time. Mean Z score fell by between 0.2 and 0.5 SDs from birth to 7 weeks, the fall tending to be greater in the earlier cohorts (table 3). The earliest dataset (Widdowson) also showed the steepest rise in Z score after the trough, so that by 52 weeks it was the heaviest, around a channel width above the median. In contrast, the second earliest dataset (CIGS) rose least and remained close to the median. Mean Z scores in the remaining datasets rose by around 0.5 SD and tended to level out after 52 weeks. It should be noted that the Widdowson and GMS datasets had only limited data in later infancy, and thus the trajectory curves are incomplete. Patterns for the girls and boys were similar (figure 1). The fitted constant SDs showed a good fit relative to the WHO standard, being close to 1 throughout. Skewness was less consistent, with negative skewness (nu < 1) in CIGS and positive skewness (nu > 1) in the other datasets. Kurtosis was near 2 as expected.

DISCUSSION
The strengths of this study are that the five datasets represent weight gain before and during the obesity epidemic. The datasets are of high quality and all but CIGS are representative of their populations, with high recruitment rates. The GAMLSS modelling approach makes maximum use of the available data and controls for potential bias. A weakness is that the most contemporary study is from Finland, not the UK, where growth patterns may differ systematically.

Overall, most cohorts were within a channel width of the WHO median and many showed an excellent fit in the early weeks, though less so beyond this point. It seems unlikely that these later differences could be simply genetic, since two of the six WHO datasets were of predominantly Northern European origin and they showed no consistent differences in stature. The lack of variation by era argues against obesity being the explanation.

The increase in birthweight over time is consistent with research in other high-income countries. This is generally thought to reflect less maternal cigarette smoking and more maternal obesity and gestational diabetes. It has been suggested that the higher birthweight seen for all these cohorts compared with the WHO standard may be explained by prior maternal undernutrition in some of the cohorts used to develop the WHO standard. However, a recent study suggested that...
birth weights across diverse countries, including some of those included in the WHO Multicentre Growth Reference Study (MGRS) cohorts, were not significantly different by country.

Though born heavier, the infants here initially lost weight relative to the WHO standard and then regained it, causing a trough to appear in the Z score growth curve. This has been described in other cohorts, and was influential in the UK rejecting the use of the WHO standard at birth. This pattern is the mirror image of the way breastfed infants used to grow on charts based mainly on bottle-fed infants and though seen in all datasets, its depth decreases over time. We do not have information about breastfeeding rates in the routine cohorts but most will have been lower than the WHO sample. We do not have information about breastfeeding rates in the routine cohorts but most will have been lower than the WHO sample. However, breastfeeding initiation rates have risen in the UK over the past 50 years from 36% in 1970 to over 70% in the early 21st century (UK) while when the Tampere cohort were born, 93% of Finnish infants were initially breastfed. The CIGS dataset, a more selective research sample, showed the shallowest trough and had higher rates of initial breastfeeding (75%) and breastfeeding to 6 months (48%). This pattern by era may also reflect changes in formula milk composition which has increasingly mirrored the nutritional content of breast milk.

Table 3  Change in mean (SD) weight Z scores between target ages by dataset

<table>
<thead>
<tr>
<th>Target age (weeks)</th>
<th>Sex</th>
<th>Change in SDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–7</td>
<td>M</td>
<td>−0.43</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>−0.40</td>
</tr>
<tr>
<td>7–52</td>
<td>M</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.06</td>
</tr>
</tbody>
</table>

CIGS, Cambridge Infant Growth Study; GDS, Growth and Development Study; GMS, Gateshead Millennium Study.
In contrast, the later rise relative to the standard did not show the same trend over time, suggesting that it may be unrelated to breastfeeding. While observational studies have found an association between use of breast milk substitutes and faster weight gain later in infancy23 1 this effect was not seen in the Belarus trial of breastfeeding promotion.33 The sampling process used to construct the WHO standard could in principle mean that the MGRS cohorts were lighter than the general population. Children who were either outliers or not breastfed to 12 months were excluded from the sample3 and it has been shown that larger children are more likely to cease breastfeeding early.13 25 However, the WHO MGRS group did not find any systematic difference in size between those included and excluded (personal communication De Onis, email communication, 2006).

Measurement error might also play a role, since the MGRS cohorts were measured using the same research protocols, where all clothes were removed or adjusted for.3 13 The present cohorts mainly used routine data collected by health professionals; UK and Finnish guidelines recommend weighing infants naked up to the age of 2 years.36 but parents of older infants may be reluctant to comply, leading to higher weights at later ages. The CIGS cohort where all infants were weighed naked to a research protocol tracks closest to the WHO standard beyond 6 months. However, another study that compared routinely collected measurements with research measurements of the same children found little difference in weight.87

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

Data sharing statement This paper uses secondary data analysis from five studies. To the authors’ knowledge none are currently publicly available.

Competing interests M012069/1.

Contributors		

Chief Scientist Of		

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.

CONCLUSIONS

The overall fit to the WHO weight standard of these cohorts ranged from excellent to adequate, so that use of the WHO standard is unlikely to introduce major bias in the assessment of individual children. Some of the more subtle variation in fit is likely to reflect variations in the levels of breast and formula feeding, as well as the composition of formula milk, in different eras. However, the lack of a consistent trend for weight gain after the initial weeks suggests that the higher average weight gain seen in North European cohorts cannot simply be explained by increasing rates of obesity in later childhood and adulthood.