Available online at www.sciencedirect.com

ScienceDirect

International Journal of Nursing Studies 46 (2009) 310–316

INTERNATIONAL JOURNAL OF NURSING STUDIES

www.elsevier.com/ijns

The reliability of routine anthropometric data collected by health workers: A cross-sectional study

William Johnson [a,*], Noël Cameron [a], Peter Dickson [b], Stuart Emsley [b,c], Pauline Raynor [c], Claire Seymour [b], John Wright [c]

[a] Centre for Human Development and Ageing, Loughborough University, UK
[b] Bradford and Airedale Teaching Primary Care Trust, Douglas Mill, UK
[c] Bradford Institute for Health Research, Bradford Royal Infirmary, UK

Received 4 June 2008; received in revised form 6 October 2008; accepted 6 October 2008

Abstract

Background: Reliable data on child growth is a prerequisite for monitoring and improving child health. Despite the extensive resources invested in recording anthropometry, there has been little research into the reliability of these data. If these measurements are unreliable growth may be misreported, and health problems may go undetected.
Objectives: To assess the reliability of routine infant growth data, following anthropometric training of health workers responsible for collecting these data, in Bradford, UK. To determine whether being observed by an external administrator influenced reliability.
Design: A test–retest design was used.
Participants: All health workers ($n = 192$) responsible for growth monitoring in Bradford were included in the study, of which 36.5% ($n = 70$) had complete data.
Methods: Following training in basic anthropometry all health workers were asked to complete a test–retest study, using infants aged 0–2 years. Health workers took two recordings of weight, length, head circumference, and abdominal circumferences on five infants. A peer health worker recorded a third set of measurements on each infant. Twenty-two individuals were selected to be observed by an external administrator during data collection. Technical error of measurements (TEMs) were produced to assess intra-observer and inter-observer reliability. Differences between groups were tested to determine whether external observation influences reliability.
Results: None of the TEMs were excessively large, and coefficients of reliability ranged from 0.96 to 1.00. All intra-observer and inter-observer TEMs for the observed group were larger than those for the non-observed group. For example, the observed group's intra-observer TEMs for weight, length, abdominal circumference, and head circumference (46.18 g, 0.60 cm, 0.65 cm, 0.47 cm) were larger than the non-observed group's TEMS (9.14 g, 0.35 cm, 0.34 cm, 0.19 cm). TEMs for weight, abdominal circumference, and head circumference were significantly larger for the observed group, compared to the non-observed group ($p < 0.001$). Inter-observer TEMs for length were also significantly larger for the observed group ($p = 0.031$), whilst intra-observer TEMs for length were not significantly different between the two groups ($p = 0.137$).
Conclusions: Following training in anthropometry health workers in Bradford can, in general, reliably measure child growth. TEMs were comparable to data from other research studies and all coefficients of reliability were indicative of good quality control. Reliability measurement provides a method of quality assurance for routine data monitoring. If commissioners of health

* Corresponding author. Tel.: +44 1509 228 159.
E-mail address: W.O.Johnson@lboro.ac.uk (W. Johnson).

0020-7489/$ – see front matter
doi:10.1016/j.ijnurstu.2008.10.003

services are to be informed by these data then some form of reliability assessment should be considered, and if employed external observation is recommended to improve validity.

Keywords: External observation; Growth monitoring; Quality assurance; Reliability; Technical error of measurement

What is already known about the topic?

- Extensive resources are invested in recording anthropometry in the UK for the purposes of growth monitoring.
- Large measurement error can influence interpretation, and in this setting may result in health problems going undetected.

What this paper adds

- Age specific TEMs from routine growth data, after initial training of health workers responsible for data collection, are comparable to those from other research projects, and should be deemed reliable.
- There is a significant difference between TEMs from health workers who self-reported reliability data and those who were observed.

1. Introduction

The growth rate of an individual is an important indicator of general health (Cameron, 2007), and growth monitoring is therefore an important screening tool. Health professionals accept growth monitoring as a standard component of community paediatric services throughout the world (Hall, 1996). Monthly growth monitoring of all children under 18 months of age is recommended in developing countries (UNICEF, 1990), and protocols with less frequent measuring are endorsed in developed countries. In the UK, growth monitoring is a standard component of child health services in the UK (Department of Health and Social Security, 1974), which involves children being regularly measured, these data being plotted on growth reference charts, and where growth is unfavourable, referral to an appropriate specialist (Garner et al., 2000). Growth monitoring distinguishes between those who demonstrate favourable growth and those who do not.

Growth monitoring has become increasingly important in the light of the epidemic of childhood obesity, which has been described as one of the most daunting public health threats facing developed countries (Department of Health, 2003). In the UK there was a statistically significant increase in body mass index (BMI) of children under 4 years of age between 1989 and 1998 (Bundred et al., 2001). Overweight and obesity are risk factors for insulin resistance and the development of the vascular and metabolic dysfunctions that precede overt cardiovascular disease and Type II diabetes (Cooper-Dehoff and Pepine, 2007). Monitoring for unfavourable growth patterns during infancy may be an important component of intervention programmes to target those at risk for obesity and non-communicable diseases later in life (Summerbell et al., 2005).

The use of Personal Child Health Records (PCHR) is endorsed in the National Service Framework for Children (Department of Health, 2004), and since 1991 PCHRs have been issued to all mothers in the UK (Wright and Reynolds, 2006). The record was developed to improve communication, enhance continuity of care, and increase parental understanding of their child's health and development (Hall and Elliman, 2003). Retention rates for the PCHR have been reported to be high throughout the UK (Hall and Elliman, 2003). Walton et al. (2006) reported that 93% ($n = 15{,}733$) of mothers, enrolled in the Millennium Cohort Study, were able to produce their PCHR when asked to by an interviewer. One main purpose of the PCHR is to provide an impetus for monitoring growth during infancy. The PCHR contains tables and charts which allow measurements of weight, length, and head circumference to be recorded throughout infancy. A national standard PCHR has been designed (Royal College of Paediatrics and Child Health, 2004), although each PCT has the choice to include pages that allow the recording of measurements before 28 days of age, at 6–8 weeks, and at 7–9 months. Therefore, there is no national growth monitoring program throughout the UK. Measurement of all children at the three defined age periods is generally recognised as routine practice, and most Primary Care Trusts (PCTs) including Bradford and Airedale PCT follow this measurement schedule (Patterson et al., 2006).

Over 90% of PCTs use health visitors to collect growth data for the PCHRs (Patterson et al., 2006), although staff nurses, community nursery nurses, and student health visitors also aid data collection. In general, staff nurses will have undergone 3 years of training to receive either a bachelors degree, an advanced diploma, or a registered general nurse qualification; and community nursery nurses will have either a BTEC National Diploma in Child Studies, or an NNEB (nursery nurse qualification) or equivalent. While health visitors will have studied for either a bachelors or postgraduate degree, and will be registered on the Nursing and Midwifery Council. Hereafter the term ‘health worker’ will be used to describe all professionals responsible for growth monitoring. The utility of the data health workers collect is dependent on its reliability. However, despite the extensive resources invested in recording growth measurements in the UK, there has been little research into reliability. This paper reports the results of a study to investigate the reliability of routine growth data as part of a cohort study.

Born in Bradford is a longitudinal birth cohort study that will investigate the determinants of childhood and adult disease. Twenty-five percent of Bradford's total population of 380,000 people are of South Asian origin. Half of the estimated 6000 annual births at Bradford Royal Infirmary (BRI) are to South Asian parents. A high proportion of these babies live in the most deprived areas of Bradford, as measured by the Index of Multiple Deprivation for England and Wales (Bradford Health Informatics Service, 2008), and associations between multiple deprivation and infant mortality in Bradford have been made (Macfarlane, 2008). In 2002, the city's infant mortality rate of 9.1 was significantly larger than the English and Welsh combined value of 5.3 (Bradford and District Infant Mortality Commission, 2005). Bradford provides the study with a unique multi-ethnic population that is characterised by high rates of infant and childhood morbidity. Growth monitoring in Bradford is fundamental to the early identification of health abnormalities, although the process of growth monitoring is not unique to Bradford. Health workers throughout the UK collect routine growth data during infancy. The initial phase of Born in Bradford aims to utilise growth data from routine health assessments. Measuring the magnitude of measurement error will help determine if these data are reliable, and is a major element of quality control (Goto and Mascie-Taylor, 2007). If these data are to be used, either for scientific purposes or to inform health service policies and recommendations, reliability must be quantified.

2. Aims

To assess the reliability of routine infant growth data for weight, height, head circumference, and abdominal circumference, following initial training of health workers responsible for collecting these data, in Bradford. To determine whether being observed by an external administrator during data collection influenced health worker's reliability.

3. Methods

All Community Practice Teachers (CPTs) and one health worker from each health centre in Bradford were invited to attend a growth-training workshop (CPTs are senior health visitors who train student health visitors during their community placement and mentor newly qualified staff). These workshops, organised by Born in Bradford in collaboration with the Child Growth Foundation, provided training on how to reliably measure weight, length, head circumference, and abdominal circumference. Supporting measurement protocols were also produced and disseminated. Health workers who attended these sessions then organised their own training days where all information was made available for their peers. At least one member of the Born in Bradford team attended all health centre training days.

Following training, all health workers in Bradford were asked to complete a test–retest reliability study. This involved taking anthropometric measurements on five infants aged less than 2 years. Discussions with Bradford and Airedale teaching Primary Care Trust (tPCT) concluded that health workers could feasibly collect data on a maximum of five infants. Each infant had three sets of measurements recorded, two by the health worker and the third by a peer health worker. Each health worker was provided with a form on which to record these data.

Measurements included weight, length, abdominal circumference, and head circumference. Infants were weighed naked, and to the last completed 10 g, using Seca baby scales. If an infant became restless weight was ascertained by measuring mother and infant together, and then subtracting the mother's weight. Length was measured to the last completed 0.5 cm using a standard issue neonatometer (Harlow Health Care, London, UK). Health workers had the choice between three pieces of equipment (Harlow Health Care, London, UK) to measure head and abdominal circumferences, all of which measured to a precision of 0.1 cm. Lassos were provided to record head circumference and abdominal circumference. Some health workers had a preference for more traditional tape measures, and these data were included in analyses.

One health worker from each health centre was randomly selected to be observed by a study administrator when collecting their data. A study administrator organised to visit these selected health workers at baby clinics, which all health centres in Bradford hold weekly (baby clinics are sessions where mothers can seek the advice of health workers, and have their babies measured and immunised). The study administrator was instructed to simply observe health workers whilst they collected their test–retest data. The study administrator ensured that each health worker understood what was asked of them, but apart from this had no other contact with the health worker during data collection. Following data collection any questions regarding the study were answered. Forms were returned by hand or via post to the study administrator, and could be divided between two groups of health workers, observed and non-observed.

The resulting data were used to produce technical error of measurements (TEMs). The TEM is the standard deviation of differences between repeated measures, uncorrelated for bias (Mueller and Martorell, 1988). In practice, this means that 95% of repeat results will fall within $\pm 1.96 \times$ TEM. In the test–retest study the differences between the first two measurements were used to produce individual intra-observer TEMs for each measurement. Similarly, the differences between the first and third measurements were used to calculate health workers inter-observer TEMs. Therefore,

Table 1
Intra-observer technical error of measurements (TEMs).

	Weight (g)	Length (cm)	Abdominal circumference (cm)	Head circumference (cm)
Total (70)				
Mean TEM	20.78	0.43	0.44	0.28
Standard deviation	50.28	0.55	0.30	0.32
Coefficient of reliability	1.00	1.00	0.99	0.99
Observed (22)				
Mean TEM	46.18	0.60	0.65	0.47
Standard deviation	72.20	0.89	0.31	0.46
Coefficient of reliability	1.00	1.00	0.97	0.98
Non-observed (48)				
Mean TEM	9.14	0.35	0.34	0.19
Standard deviation	30.69	0.26	0.25	0.16
Coefficient of reliability	1.00	1.00	1.00	1.00

for each health worker there were eight TEMs, four intra-observer and four inter-observer. Mean TEMs were calculated for the observed and non-observed groups, and for the whole sample (see Tables 1 and 2). The majority of variables were not normally distributed, and demonstrated significant positive skewing. Therefore, Mann–Whitney tests were performed to check for statistical significance between observed and non-observed data (see Tables 3 and 4).

Ethical approval for the study was granted by Bradford Research Ethics Committee, and research governance approval was provided by Bradford NHS Teaching Hospitals Trust and Bradford and Airedale PCT.

4. Results

Of the 192 health workers in Bradford 44.3% ($n = 85$) returned forms and 36.5% ($n = 70$) of these had complete data. Twenty-two health workers were observed during data collection, and 48 were not.

None of the TEMs were excessively large, and coefficients of reliability ranged from 0.96 to 1.00 (see Tables 1 and 2). Measurement error was generally higher for abdominal circumference, followed by length, and then head circumference. For example, the mean intra-observer TEM for all health workers was 0.44 cm for abdominal

Table 2
Inter-observer technical error of measurements (TEMs).

	Weight (g)	Length (cm)	Abdominal circumference (cm)	Head circumference (cm)
Total (70)				
Mean TEM	21.19	0.56	0.61	0.37
Standard deviation	50.11	0.35	0.35	0.28
Coefficient of reliability	1.00	1.00	0.98	0.99
Observed (22)				
Mean TEM	43.24	0.66	0.81	0.60
Standard deviation	72.82	0.29	0.31	0.28
Coefficient of reliability	1.00	0.99	0.96	0.97
Non-observed (48)				
Mean TEM	11.08	0.51	0.52	0.27
Standard deviation	31.39	0.37	0.32	0.21
Coefficient of reliability	1.00	1.00	0.99	1.00

Table 3
Mann–Whitney test for Intra-observer TEM data.

	Weight (g)	Length (cm)	Abdominal circumference (cm)	Head circumference (cm)
Observed (22): Median TEM (range)	14.49 (233.78)	0.37 (4.43)	0.62 (1.35)	0.33 (2.16)
Non-observed (48): Median TEM (range)	0.00 (208.71)	0.32 (1.30)	0.30 (1.13)	0.16 (0.72)
Mann–Whitney U	184.5	411.0	204.0	192.0
p value	<0.001**	0.137	<0.001**	<0.001**

** Significant at alpha 1% level.

Table 4
Mann–Whitney test for Inter-observer TEM data.

	Weight (g)	Length (cm)	Abdominal circumference (cm)	Head circumference (cm)
Observed (22): Median TEM (range)	9.85 (229.26)	0.67 (1.33)	0.88 (1.18)	0.51 (1.21)
Non-observed (48): Median TEM (range)	0.00 (208.71)	0.46 (1.86)	0.49 (1.45)	0.21 (1.25)
Mann–Whitney U	253.50	358.00	264.50	101.50
p value	<0.001**	0.031*	0.001**	<0.001**

* Significant at alpha 5% level.
** Significant at alpha 1% level.

circumference, compared to 0.28 cm for head circumference. In practice this means that 95% of repeat measures for abdominal circumference and head circumference will fall within ±0.86 cm and ±0.55 cm, respectively (i.e. 1.96× TEM). All inter-observer TEMs, apart from weight in the observed group, were larger than the respective intra-observer TEMs. Measurement error was larger in the observed group, and this is reflected by larger TEMs. For example, the observed group's intra-observer TEMs for weight, length, abdominal circumference, and head circumference (46.18 g, 0.60 cm, 0.65 cm, 0.47 cm) were larger than the non-observed group's TEMS (9.14 g, 0.35 cm, 0.34 cm, 0.19 cm). This pattern was present for both intra-observer and inter-observer data.

There were significant differences between the observed and the non-observed groups' TEMs (see Tables 3 and 4). Generally, measurement error was significantly higher in the observed group, compared to the non-observed group. Intra-observer TEMs for weight, abdominal circumference, and head circumference were significantly larger for the observed group, compared to TEMs for the non-observed group ($p < 0.001$). Intra-observer TEMs for length were not significantly different between the two groups ($p = 0.137$). Similarly, all inter-observer TEMs were significantly larger for the observed group at alpha 1% ($p < 0.001$), apart from length which was significantly larger at alpha 5% ($p = 0.031$).

5. Discussion

After training in basic anthropometry, TEMs from health workers in Bradford were comparable to published TEMs from research studies that reported acceptable levels of reliability (Ulijaszek and Kerr, 1999). A general conclusion that health workers can reliably measure child growth can be made. However, health workers who were observed by a study administrator during data collection had higher levels of measurement error than those who were not observed.

All health workers responsible for growth monitoring in the community were included in the study, making the total sample externally valid. Only 36.5% of health workers returned forms with complete data. However, we have no reason to believe that these individuals differed in any way from the total health worker population (e.g. sex, level of education attained, and duration of employment as a health worker). Complete data was collected on health workers from different geographical locations across Bradford. Our sample does not, therefore, neglect health workers who monitor growth in areas of Bradford which have important defining characteristics. For example, areas with high levels of deprivation, or areas that are predominantly occupied by South Asian populations. For these reasons we believe our total sample is representative of all health workers in Bradford. The study administrator aimed to observe one health worker from each health centre although time constraints did not allow this. Individuals in the observed group were selected at random and were likely to be representative of health workers with varying levels of enthusiasm to participate in the study. Whereas, it is likely that only the most enthusiastic health workers returned forms in the non-observed group. There may, therefore, be selection bias in the non-observed group. If the reliability of all health workers was routinely assessed data from the less enthusiastic individuals could be collected.

A paucity of research reporting age-specific TEMs meant that power calculations could not be performed. Whilst our total sample is representative of health workers in Bradford, a larger sample size would have further increased the power to detect statistically significant differences in TEMs. This study only assessed the reliability of health workers in Bradford, and no comparable age-specific TEMs from health workers in other cities or counties have been published. It is important to reiterate that health workers and their involvement in growth monitoring are not unique to Bradford. Health workers with similar levels of education, training, and experience measure infant growth at routine age periods in other cities and counties, and for this reason we would expect similar levels of reliability throughout the UK.

The large number of health workers in health services responsible for collecting anthropometric data increases the likelihood that one person's measurements will differ significantly from another's (Ulijaszek and Kerr, 1999). The difference between repeat measurements has been termed measurement error, and in this context has been used to explain the extent to which repeat measures give the same value (Habicht et al., 1979). Large measurement error can influence interpretation and limit the usefulness of growth data (Ulijaszek and Kerr, 1999). Growth monitoring is used to assess the growth of an individual between two, or more time points, and thus depends on a series of recordings. Small measurement error for any one recording is unlikely to have clinical significance, but systematic measurement error for two or more recordings will decrease the ability of growth monitoring to identify failure to thrive. The measurement error of routine growth data has clinical importance which, in part, determines the validity of growth monitoring.

The TEM is the statistic most commonly used to explain measurement error (Mueller and Martorell, 1988), and can provide sufficient information to determine whether a set of anthropometric measurements are reliable. The coefficient of reliability (*r*) reveals what proportion of variance is free from measurement error. Coefficients of reliability above 0.95 are indicative of good quality control (Goto and Mascie-Taylor, 2007).

The TEMs from this study are similar to acceptable levels of reliability found in anthropometric literature, and all coefficients of reliability were above 0.95. For these reasons, our TEMs indicate good reliability of growth measures. Compared to the mean TEMs reported in Ulijaszek and Kerr's (1999) review our TEMs for weight and abdominal circumference were smaller. This is surprising considering that Ulijaszek and Kerr conducted a review of research studies, where data was collected by trained anthropometrists. Our intra-observer TEM for length (0.43 cm) was within the range (0.10–0.80 cm) reported in Ulijaszek and Kerr's (1999) review, and our inter-observer TEM for length (0.56 cm) was just outside the range (0.1–0.5 cm). Compared against reliability data (WHO Multicentre Growth Reference Study Group, 2006), on anthropometrists trained to measure infants in the WHO Multicentre Growth Reference Study (MGRS), our TEMs for length and head circumference were similar.

The mean TEMs reported in Ulijaszek and Kerr's (1999) review included results from data on infants, children, and adults. As the absolute measurement increases it is likely that absolute measurement error also increases. It could, therefore, be assumed that TEMs from data on adults would be larger than TEMs from data on infants. This may be why our TEMs for weight and abdominal circumference are smaller than those reported by Ulijaszek and Kerr. There may be a need for age specific TEMs. However, It is unlikely that the magnitude of the measurement will affect reliability within our age range (0–2 years). The MGRS have reported age specific TEMs for infants aged 0–24 h, and another set for infants aged 0–1 years (WHO Multicentre Growth Reference Study Group, 2006). Our TEMs for length and head circumference are comparable with these data. Other studies have reported age specific (1–2 years) intra-observer and inter-observer TEMs for length of 0.4 and 0.5 cm, respectively (Ulijaszek, unpublished; Pelletier et al., 1991). Our mean TEMs for all health workers were almost identical to these data (0.43 and 0.56 cm). We are not aware of published age specific TEMs for weight and abdominal circumference during infancy.

In general, the inter-observer TEMs from this study were marginally larger than the intra-observer TEMs. It might also be expected that the difference between two recordings taken by the same person should be smaller than the difference if two people took one recording each. However, it is far from universally the case that intra-observers TEMs are smaller then inter-observer TEMs (Ulijaszek and Lourie, 1994). Using data from the Malawi Maternal and Child Health Survey, Pelletier et al. (1991) found intra-observer error to be greater than inter-observer error for length and arm circumference. Larger intra-observer errors have also been reported for subscapular skinfolds in a United States population (Johnston et al., 1972).

The observed group's TEMS were, in general, significantly larger than the non-observed group's. There are a number of possible reasons for this. Firstly, the presence of an observer distracted or intimidated health workers resulting in larger TEMS. Secondly, health workers in the non-observed group felt like they were being judged and reported more favourable results to appear more reliable. Throughout the study health workers were assured that variability is an inherent part of the measurement process. However, health workers had never been asked to complete a reliability study before and may have felt expectations to report high reliability. Health workers in the non-observed group were more likely to report both their first and second recording, for a measurement, to be the same. There were also more occurrences in the non-observed data where all three recordings were the same. Also, health workers in the non-observed group reported head and abdominal circumferences to the nearest 0.5 cm more frequently than health workers in the observed group. If this is because of rounding results up/down health workers in the non-observed group did not measure to the full precision of the instruments. For these reasons we believe that self-reported reliability checks may produce favourable results, hence TEMs for the non-observed group should be interpreted with caution. The results of this study should be used to emphasise the normal variation expected between repeat measurements in future documentation and training of anthropometry.

TEMs from routine growth data collected by health workers in Bradford indicate acceptable levels of measurement error. TEMs were calculated from data collected by health workers, after they had been trained in basic anthropometry. This was, in effect, an intervention study, and reliability after training is acceptable. Training in anthropometry and the production of a measurement protocol may have helped to standardise measurement technique of health workers in Bradford, improving reliability. Although, without test–retest data available prior to training this hypothesis cannot be tested.

Extensive resources are invested in collecting and recording growth measurements in developed and developing countries throughout the world. In the UK, there has been no research into the reliability of these measurements. Routine growth monitoring produces an unexploited source of data for public health surveillance, and our results suggest that with initial training in measurement techniques these data can be of research calibre.

Health care commissioners require accurate growth data if they are going to make evidence based decisions on local policy and provision of services. Reliability checks, including external observation, of intra-observer and inter-observer error should be considered to measure the accuracy of growth data. As well as measuring accuracy, reliability checks reinforce the importance of standards and act as a quality assurance mechanism with feedback to practitioners.

Acknowledgements

This research was generously supported by the Child Growth Foundation and could not have been completed without the time and co-operation of the health visitors of Bradford who have our gratitude.

Conflicts of interest

None.

Funding

Child Growth Foundation.

Ethical approval

Ethical approval for the study was granted by Bradford Research Ethics Committee (07/Q1202/38), and research governance approval was provided by Bradford NHS Teaching Hospitals Trust and Bradford and Airedale PCT.

References

Bradford and District Infant Mortality Commission, 2005. Interim Report on Progress.

Bradford Health Informatics Service, 2008. Analysis and Interpretation of Infant Mortality Data. Available: www.bdimc.bradford.nhs.uk (accessed on 05.06.2008).

Bundred, P., Kitchiner, D., Buchan, I., 2001. Prevalence of overweight and obese children between 1989 and 1998: population based series of cross sectional studies. British Medical Journal 322 (7282), 326–328.

Cameron, N., 2007. Growth patterns in adverse environments. American Journal of Human Biology 19 (5), 615–621.

Cooper-Dehoff, R.M., Pepine, C.J., 2007. Metabolic syndrome and cardiovascular disease: challenges and opportunities. Clinical Cardiology 30 (12), 593–597.

Department of Health, 2003. Annual Report of the Chief Medical Officer 2002. Health Check: On the State of the Public Health. Department of Health Publications, London.

Department of Health, 2004. Care Document, National Service Framework for Children, Young People and Maternity Services. Department of Health Publications, London.

Department of Health and Social Security, 1974. Present Day Practice in Infant Feeding. HMSO, London.

Garner, P., Panpanich, R., Logan, S., 2000. Is routine growth monitoring effective? A systematic review of trials. Archives of Disease in Childhood 82 (3), 197–201.

Goto, R., Mascie-Taylor, C.G., 2007. Precision of measurement as a component of human variation. Journal of Physiological Anthropology 26 (2), 253–256.

Habicht, J.P., Yarbrough, C., Martorell, R., 1979. Anthropometric field methods: criteria for selection. In: Jelliffe, D.B., Jelliffee, E.F.B. (Eds.), Nutrition and Growth. Plenum Press, New York, pp. 365–387.

Hall, D.M.B., 1996. Health for All Children. Report on the Third Joint Working Party on Child Health Surveillance. Oxford University Press, Oxford.

Hall, D.M.B., Elliman, D., 2003. Health for All Children, 4th ed. Oxford University Press, New York.

Johnston, F.E., Hamill, P.V., Lemeshow, S., 1972. Skinfold thickness of children 6-11 years, United States. Vital and Health Statistics. Series 11, Data from the National Health Survey 120(120), 1–60.

Macfarlane, A., 2008. Trends and Variations in Births and Infant Mortality in England and Wales. Available: www.bdimc.bradford.nhs.uk (accessed on 05.06.2008).

Mueller, W.H., Martorell, R., 1988. Reliability and accuracy of measurement. In: Lohman, T.G., Roche, A.F., Martorell, R. (Eds.), Anthropometric Standardisation Reference Manual. Human Kinetics Books, Illinois, pp. 83–86.

Patterson, L., Jarvis, P., Verma, A., Harrison, R., Buchan, I., 2006. Measuring children and monitoring obesity: surveys of English primary care trusts 2004–2006. Journal of Public Health 28 (4), 330–336.

Pelletier, D.L., Low, J.W., Msukwa, L.A.H., 1991. Sources of measurement variation in child anthropometry in the Malawi maternal and child nutrition survey. American Journal of Human Biology 3, 227–237.

Royal College of Paediatrics and Child Health, 2004. The National Standard PCHR. Available: http://www.health-for-all-children.co.uk/?SHOP=HFAC4&DO=USERPAGE&PAGE=PCHRPDF (accessed on 13.08.2008).

Summerbell, C.D., Waters, E., Edmunds, L.D., Kelly, S., Brown, T., Campbell, K.J., 2005. Interventions for Preventing Obesity in Children. Cochrane Database of Systematic Reviews (Online) 3(3), CD001871.

Ulijaszek, S.J., Unpublished data.

Ulijaszek, S.J., Kerr, D.A., 1999. Anthropometric measurement error and the assessment of nutritional status. The British Journal of Nutrition 82 (3), 165–177.

Ulijaszek, S.J., Lourie, J.A., 1994. Intra- and inter-observer error in anthropometric measurement. In: Ulijaszek, S.J., Mascie-Taylor, C.G.N. (Eds.), Anthropometry: The Individual and the Population. Cambridge University Press, Cambridge, pp. 30–56.

UNICEF, 1990. Strategy for Improved Nutrition of Children and Women in Developing Countries. United Nations Children's Fund, New York.

Walton, S., Bedford, H., Dezateux, C., Millennium Cohort Study Child Health Group, 2006. Use of personal child health records in the UK: findings from the millennium cohort study. British Medical Journal 332 (7536), 269–270.

WHO Multicentre Growth Reference Study Group, 2006. Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study. Acta Paediatrica 450, 38–46.

Wright, C.M., Reynolds, L., 2006. How widely are personal child health records used and are they effective health education tools? A comparison of two records. Child: Care, Health and Development 32 (1), 55–61.