A multicentre randomised controlled trial of an intervention to improve the accuracy of linear growth measurement

T H Lipman, K D Hench, T Benyi, J Delaune, K A Gilluly, L Johnson, M G Johnson, H McKnight-Menci, D Shorkey, J Shults, F L Waite, C Weber

....

See end of article for authors' affiliations
....

Correspondence to:
Associate Professor
T H Lipman, University of Pennsylvania, School of Nursing, 420 Guardian Drive, Philadelphia, PA 19104-6096, USA;
lipman@nursing.upenn.edu

Aims: To evaluate linear growth assessment and the effect of an intervention on measurement accuracy in primary care practices (PCP) within eight US geographical areas.

Methods: In this multicentre randomised controlled intervention study, paediatric endocrine nurses as site coordinators (SC) visited 55 randomly assigned PCP to evaluate growth assessment of staff performing linear measurements. SC observed 127 measurers assessing a total of 878 children: 307 (baseline), 282 (3 months), and 289 (6 months). Accuracy was determined by SC re-measuring each child with correct technique and equipment. State of the art equipment and a standardised growth training session were provided to the intervention group (IG) following the baseline visit. SC repeated data collection at all PCP at 3 and 6 months.

Results: There were no baseline differences between IG and CG equipment, technique, or accuracy; only 30% of measurements were accurate ($\leqslant 0.5$ cm from SC). Post-intervention, significantly more IG measurements were accurate: IG = 55%, CG = 37% at 3 months; IG = 70%, CG = 34% at 6 months. Odds ratio of accuracy for IG versus CG was 2.1 at 3 months and 4.5 at 6 months. At 6 months, mean difference from the SC measurements was 0.5 cm in IG and 1.1 cm in CG.

Conclusions: In PCP, children are measured inaccurately. Our intervention significantly improved measurement accuracy. Improved accuracy could yield more rapid detection and diagnosis of paediatric growth disorders.

Growth is the single most important indication of the health of a child.[1] Benefits of growth monitoring include identification of chronic disorders and provision of reassurance to parents.[2] Over 65 years ago, Meredith stated that any research involving the physical growth of a child must employ unusually rigorous measurement technique.[3] Data still demonstrate that children are frequently measured incorrectly.[4–6] To be considered accurate, a single height measurement must be within 0.5 cm (the 95% confidence limit) of the "true" height.[4] Measurements may be inaccurate by considerably more because of faulty equipment,[4] the use of incorrect technique by the auxologist,[6] incorrect criteria for obtaining length versus height,[5] or movement of the child.[4] The error is even greater when children younger than age 2 are measured.[7 8] Since children 2 years old to the age of puberty grow an average of only 5 cm a year, measurement inaccuracies may have a significant impact on the diagnosis of growth disorders. Growth failure may be misdiagnosed and underlying disorders could go unrecognised.

When obtaining height, children should be measured while standing against a wall mounted device with a fixed right angle at the head. Children younger than 24 months must be measured supine on a firm platform with a yardstick attached, a fixed head plate, and a moveable footplate.[9] Due to slight changes in a child's posture with each measurement, it is recommended that three consecutive linear measurements be obtained on each child. The average of the three measurements is considered to be closer to the true height of the child.[4] Many examiners do not employ the correct technique for obtaining linear measurements. Waite[6] observed 30 children being measured, of which 16 were measured with incorrect technique; babies were measured in their mothers' arms and children were measured wearing

their shoes. In a survey of 50 primary care practices (24 family practices, 26 paediatric practices), 58% had an incorrect policy regarding when to obtain length versus height.[5] Measurements obtained by different observers were significantly different (p < 0.001) when children aged 18–24 months were measured.[7] Measuring instruments must be accurate and installed accurately. In a study reviewing measuring instruments in over 200 health centres, general practices, hospital clinics, and schools, one in seven was found to be inaccurate by at least 1 cm and a few by considerably more.[4] When assessing growth in a multicentre study, the recommendation is to evaluate the accuracy of the measuring device used at each site[10] or provide each site with the same measuring device installed accurately.[11] Accurate equipment need not be expensive. Voss and Bailey[12] demonstrated that inexpensive height measuring equipment, once accurately installed, is no less reliable than the most expensive.

The purpose of this randomised controlled multicentre study was to determine whether instructing health care providers on correct measurement technique and the use of accurate equipment results in more accurate linear measurements of children.

METHODS
Study design
Study staff included the site coordinators (SC) and the principal investigators (PI). SC were paediatric endocrine nurses with an average of eight years of experience

....

Abbreviations: CG, control group; IG, intervention group; GEE, generalised estimating equation; PCP, primary care practice; PI, principal investigator; SC, site coordinator; TS, training session

measuring children. The PI were doctorally prepared advanced practice paediatric endocrine nurses with extensive experience in growth assessment and research and were responsible for the study design, coordination, and training curricula. They set the standard for measurement accuracy.

The study was divided into two phases; the focus of this paper is phase 2. Phase 1 consisted of a mailed survey on linear growth practices sent to a randomly selected group of 1300 paediatric and family primary care practices (PCP) in eight areas of the United States: Philadelphia, PA; Galveston, TX; St Louis, MO; New Orleans, LA; Providence, RI; Broward County, FL; Albany, NY; and Tampa Bay, FL. These areas were the cities of residence for the SC. A 19.8% response rate was achieved.

In phase 2 of the study, from the 259 respondents to the mailed survey, eight PCP per geographical site were derived using a table of random numbers. An attempt was made to secure a 3:1 representation of paediatric versus family practice providers based on the relative proportion of each who returned the survey in each geographical area during phase 1. If a selected practice declined participation or did not meet eligibility criteria, the next practice identified on the randomisation table was invited to participate. The Tampa Bay site withdrew resulting in seven geographical sites for phase 2.

At least two staff/practice who performed linear growth measurements (referred to as measurers) or approximately 16 measurers per site were recruited to achieve 89% power at a level of significance of 0.05 to detect measurer ability to obtain measurements within 0.5 cm of SC. Institutional Review Board approval was obtained from each SC's institution, as appropriate, and from the University of Pennsylvania.

Study sites

Each SC contacted selected physicians/office managers and invited the practice to participate in the six month study to evaluate linear growth. Practices were excluded if <300 children (<19 years old) were followed in the practice; if there were no measurers who were fluent in English; or if there were not at least two measurers who planned to remain on staff for at least the six month duration of the observation period. Within each geographical site, from the randomly ordered list, the six paediatric practices were alternately assigned to either the intervention group (IG) or control group (CG). The two family practices were also alternately assigned to either the IG or CG.

Intervention

At the beginning of phase 2, each SC attended a two day training session (TS) that reviewed the study protocol and a detailed two hour standardised TS that each SC would provide to the measurers in the IG. Practice sessions on performing recumbent length measurements on infants (<2 years of age) and standing height measurements on children (≥2 years of age) were also provided. Each SC's growth measurement technique was standardised to study requirements (fig 1). Accuracy of an individual SC's measurements was assured through inter-observer reliability between PI and SC measuring the same child. Measurements were considered acceptable if the difference was ≤0.3 cm. Because the SC measurements were regarded as the ''gold standard'' at the site visits, more stringent criteria were required for inter-observer reliability between PI and SC than the 0.5 cm difference that was required between SC and measurers. Intra-observer reliability was assured, in that the SC repeated measures were required to be within 0.3 cm.
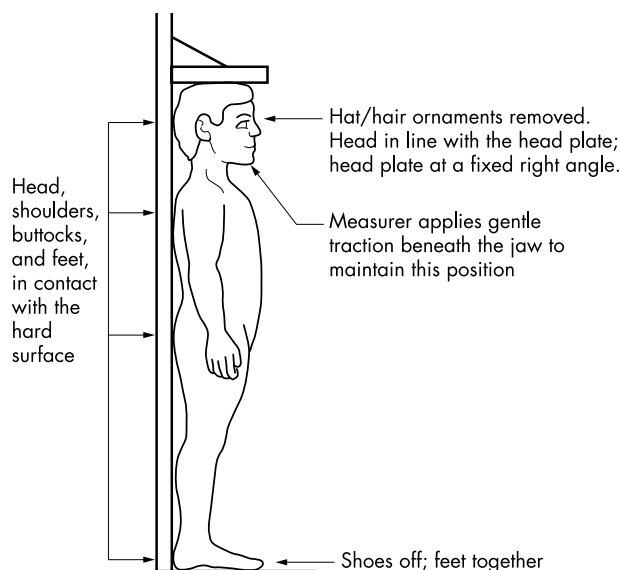

Figure 1 Illustration of correct measuring technique.

Study protocol

At the baseline visit to IG and CG practices, SC observed 2–4 measurers per practice performing measurements on at least two different paediatric patients using their usual technique and equipment. The SC documented the measurer's technique and equipment for each measurement. However, SC were blinded to the numeric value of the measurement. SC then performed three measurements on the same child using the state-of-the-art portable equipment appropriate for the child's age; the measurer was not permitted to observe these measurements. For height measurements, after the baseline visit IG were provided with a permanently affixed, wall mounted plastic accustat with a level rolling head plate (Clover Global Group, Inc., Chicago, IL, USA). For length measurements, the Pediatric Length Board (PLB) by O'Leary (Ellard Instrumentation Ltd, Seattle, WA, USA) was provided. The PLB was made of clear acrylic with a stationary 90 degree angle head piece with a removable, sliding 90 degree angle foot piece. Both measuring devices provided measurements in Imperial and metric with calibration to 1/8th inch or 1 mm respectively. The SC measurements were averaged, and the result recorded as a single measurement. The SC then obtained the numeric value of the measurements from the measurers. A single measurement was performed by the measurer because measuring each child three times was impractical in clinical practice. Both the measurers' and SC's measurements were recorded on the study's standardised observation record that included 20 descriptors of measurement technique (table 1).

After the baseline data were collected, the SC delivered the TS and the measuring equipment described above to IG measurers. The TS included a written pre-test of knowledge of appropriate linear growth measurement, a slide presentation, and handouts presenting the physiology and

---

Table 1 Summary of descriptive data collected on measurement technique

Position of child: lying versus standing
Description of measurement equipment
Presence of shoes, socks, braids, hair ornaments, hats
Position of child's head, heels, buttocks, and shoulders
Application of gentle traction to child's jaw

pathophysiology of growth disorders and linear growth monitoring techniques. The TS also included use and accurate installation of measurement equipment, a demonstration/return demonstration of correct length and height measurements, and a written post-test assessment. This equipment remained in the IG practice for staff to use on all patients. Practices participating in the CG group were not provided with the TS, equipment, or any information on growth or linear growth measurement technique. At 3 and 6 months after baseline, the SC collected follow up data on growth measurement practices using the procedure outlined for the baseline visit observation; the TS was not repeated. Inter-observer reliability was reconfirmed in each site by a PI at the time of visit 2.

Statistical analysis

Analyses were performed using STATA 6.0. Inter-observer reliability between PIs was assessed using the kappa statistic. Kappa greater than 0.75 was considered excellent agreement beyond chance. Two sided tests were conducted and a p value <0.05 was considered statistically significant.

We hypothesised that the intervention would produce more accurate measurements. The primary study outcomes were binary and the unit of analysis was the measurement of the child. Equipment was coded as ''accurate'' based on recommended criteria.[9] Measurement technique was coded as correct if all components in fig 1 were correct. Measurements were coded as ''accurate'' if the measurer's value and the site coordinator's value did not differ by greater than 0.5 cm (95% CI). The percentage of accurate measurements and measurements obtained with correct technique at each visit was first compared between the IG and CG using the χ^2 test.

The difference in measurements between SC and measurer was also compared between groups using Student's t test. To adjust for the fact that measurers in this study made multiple measurements and that multiple PCP were evaluated, a multivariate analysis was conducted using the generalised estimating equation (GEE) approach. This analysis allowed for estimation of the odds of accuracy and correct technique in the IG versus CG, while adjusting for other variables. GEE models were fit at each visit. In addition, a final model was constructed based on data from all visits. The final model thus assessed the impact of the intervention over time.

RESULTS

Enrolment

Approximately 6% of practices contacted did not meet all eligibility criteria and thus were not invited to participate. Five per cent of those invited to participate, refused. Primary reasons given by practices that refused participation included: (1) the pending or present reorganisation of the practice caused by managed care, practice consolidation, or dissolution; and (2) provider unwillingness due to low importance assigned to linear measurements. Two sites, MO and NY, had only one family practice that met eligibility criteria and agreed to participate. Both of these sites recruited an additional paediatric practice to participate in order to meet the study requirements for the total number of practices. The FL site was only able to recruit a total of seven practices. Thus a total of 55 practices (44 paediatric and 11 family practice) were assigned to the IG or CG. The group assignment was not disclosed to the practices. Practices cared for an average of 4000 children (range 300–15 000 children) and employed an average of 3.6 staff responsible for measuring children. A total of 127 measurers were assessed at baseline. By the 3 month visit, 15 measurers had left the PCP because of illness or change of jobs. The remaining 112 measurers were assessed at 3 months and 107 at 6 months.

Table 2 The sample

	Practices	Measurers	Children measured
Baseline	55 (11 FP, 44 PP)	127	307
3 months	55 (11 FP, 44 PP)	112	282
6 months	53 (10 FP, 43 PP)	107	289

Twenty one per cent of the measurers were registered nurses, 23% were licensed practical nurses, and 56% were nurses' aides/medical assistants. The measurers had an average of 8.2 years experience of measuring children. At baseline, 307 children were measured, 282 children at 3 months, and 289 at 6 months, for a total of 878 children (table 2). Children ranged in age from newborn to 18 years (mean 5.0 years); 54% were female. There were no significant differences between the demographic data of the intervention and control group practices, measurers, or children measured.

Outcomes

At baseline there were no significant differences between the IG and CG in their measuring equipment, technique, or accuracy. Table 3 shows the most commonly used measuring devices at baseline. Only 58% of children measured standing and 18% of children measured lying were measured with accurate equipment. At baseline, correct measurement technique was demonstrated in only 30% of children measured overall; 29% of children in the CG and 32% of children in the IG. Measurements by the PCP staff differed by as much as 12.1 cm from the site coordinators and were accurate ($\leqslant$0.5 cm from site coordinator) in 28% of children measured in family practices and 31% of children measured in paediatric practices. When assessing the mean difference of measurements by staff compared to the site coordinators, at baseline, the measurements differed by an average of 1.3 (1.6) cm in the CG and 1.2 (1.2) cm in the IG.

A single PCP staff measurement was compared to the mean of the SC three measurements. The mean difference between the SC first measurement and the average of the SC three measurements was −0.008 (first − average) with a 95% confidence interval of −0.02 to 0.004. The narrow width of this confidence interval around zero suggests that the two variables were extremely close in value and did not affect the results.

Significantly more children were measured with correct technique 3 months after the intervention; 54% in the IG versus 23% in the CG (p < 0.0005). At the 6 month visit, 74% of children in the IG and 26% of children in the CG were measured with correct technique (p < 0.0005) (fig 2). Accuracy of measurements was also greatly improved from a mean difference between the measurements by the PCP

Table 3 Measurement devices used in PCP at baseline

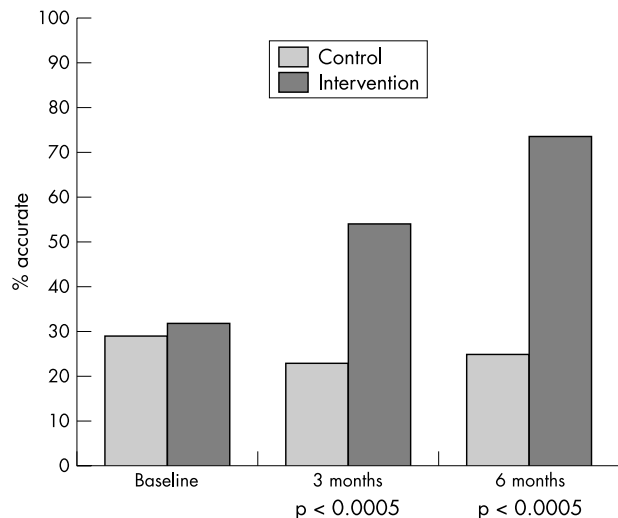|	n	%
Height devices		
Stadiometer	113	(58)
Floppy arm on scale	57	(30)
Measuring tape/stick attached to wall	23	(12)
Length devices		
Mark exam table paper and measure	55	(48)
Measuring tape alone placed on child	20	(18)
Pediatric Length Board	20	(18)
Cushioned table infant scale	11	(10)
Measuring mat	8	(7)

Figure 2 Percentage of children measured with correct technique; intervention versus control group.
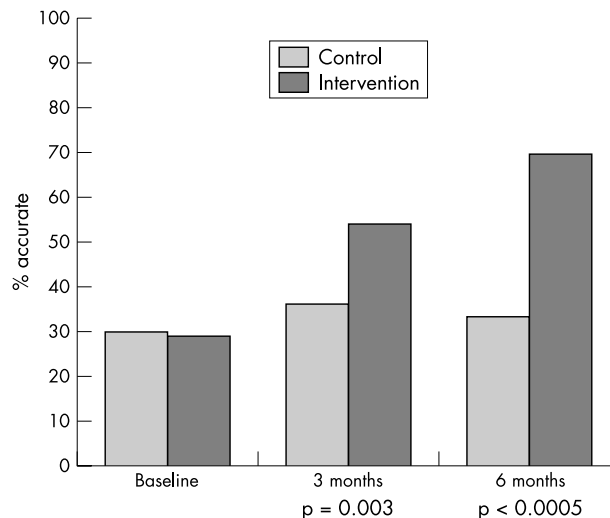

Figure 3 Percentage of children measured accurately; intervention versus control group.

staff and the site coordinators of 1.2 cm at baseline to 0.5 cm at 6 months (table 4). Post-intervention, significantly more measurements were accurate; 55% of the children measured in the IG practices, and 37% of the children measured at the CG practices were measured accurately at 3 months (p = 0.003). By 6 months, 70% of the children measured in the IG practices and 34% of the children measured in the CG practices were measured accurately (p < 0.0005) (fig 3). The intervention effect was stronger with time. The unadjusted odds ratio of accuracy for intervention group versus control group was 2.1 (3 mth) and 4.5 (6 mth).

Multivariate analysis showed that age of the child was associated with accuracy of measurement. For each five year increase in the child's age, children were 1.3 times more likely to be measured accurately (p = 0.005). Educational background of the measurer was also associated with accuracy of measurement (p = 0.002). Registered nurses were twice as likely as measurers of other educational backgrounds to be accurate at 3 months (p = 0.01). A generalised estimating equation (GEE) was used to control for age of the child and educational background of the measurer in assessing the effect of the intervention. By 6 months, the IG was twice as likely to use correct technique and five times as likely to be accurate compared to CG when controlling for child's age, measurers' education, and repeated measures (p < 0.0005, p < 0.0005). A GEE was used to assess any differences based on multiple practices. The intervention remained significant, and the estimate of the regression coefficient associated with the interaction between the intervention group and practice was within 10%.

DISCUSSION
Our intervention programme significantly improved the accuracy of linear measurements in primary care practices.

The findings from this study show that, in primary care practices, many children are measured with inaccurate equipment and incorrect technique which results in inaccurate measurements. Most of the equipment used by practitioners in this study cannot yield accurate measurements because of the inherent inaccuracy of the equipment. The floppy arm of a balance scale, used to measure the height of 30% of the children, is flawed because there is no fixed right angle at the head and no backboard to stabilise the child's posture. The equipment used to measure length is even more inaccurate. Marking examining room paper at the head and foot while the child lay on a cushioned exam table was the method used to measure the length of 48% of the children in the study. This method is extremely inaccurate because of the movement of the child and the crevices formed on the paper as the child lies on the mattress. This inexact method partially explains the greater degree of inaccuracy when younger children are measured. We did not obtain data to determine why primary care practitioners used inaccurate equipment, although we believe that the degree of inaccuracy caused by poor equipment is generally unknown.

Our finding that incorrect technique used when obtaining linear measurements is a major factor contributing to inaccuracy, has also been shown by others.[4 6 7] Measurements of 70% of the children were obtained with inaccurate technique ranging from the child wearing shoes, to the feet being splayed during measurement. Even when using state of the art equipment, incorrect technique yields inaccurate measurements. It appears the reason for failure to use correct technique was lack of knowledge. It is important to note that at 3 months, registered nurses were significantly more likely to obtain accurate measurements than measurers with other educational backgrounds. The education of registered nurses may include lectures on normal growth, growth disorders, and the importance of linear measurement,

Table 4 Effect of intervention on accuracy of linear measurement			
Group	**Baseline x̄ Diff. (SD)**	**3 months x̄ Diff. (SD)**	**6 months x̄ Diff. (SD)**
Control (n)	1.3 (1.6) cm (n = 169)	1.0 (0.9) cm (n = 155)	1.1 (1.4) cm (n = 157)
Intervention (n)	1.2 (1.2) cm (n = 138)	0.6 (0.7) cm (n = 127)	0.5 (0.8) cm (n = 132)
		p = 0.0001	p < 0.0005

which could increase their motivation to obtain measurements using correct technique.

Through our intervention programme, the accuracy of linear measurements in primary care practices was significantly improved. The average degree of inaccuracy was decreased from 1.2 cm to 0.5 cm. Prior to beginning the study, we speculated that the accuracy of children measured by the control group would also improve by our mere presence. That did not occur, however, most likely because the measurers were not aware of what they were doing incorrectly. We were also concerned that the effect of our intervention might diminish with time; that the education programme would be forgotten. The converse was true, with the effect becoming stronger with time, indicating that continued practice with correct technique and accurate equipment yielded greater improvement. In addition, there may have been some continual reinforcement of the importance of correct technique. Over time, the measurers were able to compare their own previous measurements on the same patients and observe the degree of inaccuracy.

Although our intervention programme had a marked effect on improving accuracy, 30% of the measurements obtained on children in the intervention group were still not accurate at 6 months. We hypothesise that convenience was one impediment to changing practice. A disadvantage to the infant length board is that it was not available in each examining room and the measurers had to carry the board into each room. The method of marking the exam paper, although extremely inaccurate, has the advantage of being readily available in each examining room. Another issue is that, in general, those who set policy within the practices did not mandate a change in measuring practice. Change in practice is much more likely to occur if the need for accountability is clearly defined. We purposely did not direct our education programme to the physicians and practice administrators, wanting to educate those who were responsible for measuring children. That is, however, a limitation of this study. After the data were analysed we provided the physicians with aggregate data related to the degree of measurement inaccuracy we demonstrated. We must consider that physicians may not be knowledgeable about the clinical significance of measurement inaccuracy. We demonstrated that measurements were inaccurate by an average of 1.3 cm at the baseline visit. Normal childhood growth is 5 cm per year and the definition of growth failure is growth less than 4 cm per year.[13] An inaccurate measurement of 1.3 cm could cause the lack of detection of growth failure or the inappropriate referral of a normally growing child. The clinical significance of linear growth failure is great as a previously undiagnosed organic disease can be identified. In a study of over 14 000 school children in the United Kingdom, 140 previously undiagnosed children had heights below the 3rd centile. Of these children, 23% had organic disease identified based on the height screening. The diseases included, but were not limited to, Noonan's syndrome, coeliac disease, hypothyroidism, and growth hormone deficiency.[14] Of 198 children referred to an Oxford growth clinic because of short stature, 40% had an organic reason for abnormal growth.[15] In view of the ability to diagnose remediable diseases through accurate growth assessment, the clinical significance is clear.

These data have definite implications for policy and standards of practice related to children's health care. Registered nurses were twice as likely to measure children accurately. This provides additional evidence demonstrating the improved quality of care delivered by registered nurses and should be considered when hiring staff to work in paediatric practices. As shown in this study, child health policy guidelines related to practice must mandate that paediatric health care providers obtain accurate measuring equipment and ensure that staff are adequately trained on correct measurement technique by experts. This intervention was successful because experienced endocrine nurses were the trainers. The ultimate goal is to obtain accurate linear measurements optimising growth assessment for all children.

. . . .

Authors' affiliations
T H Lipman, K D Hench, T Benyi, J Delaune, K A Gilluly, L Johnson, M G Johnson, D Shorkey, F L Waite, C Weber, Pediatric Endocrinology Nursing Society Multicenter Study, Gaithersburg, MD, USA
T H Lipman, H McKnight-Menci, J Shults, University of Pennsylvania, Philadelphia, PA, USA

REFERENCES
1 **Tanner JM**. Normal growth and techniques of growth assessment. *Clin Endocrinol Metab* 1986;**15**:411–51.
2 **Hall D**. Growth monitoring. *Arch Dis Child* 2000;**82**:10–15.
3 **Meredith HV**. The reliability of anthropometric measurements taken on eight and nine year old white males. *Child Dev* 1936;**7**:262–72.
4 **Voss LD**, Bailey BJR, Cumming K, *et al*. The reliability of height measurement (the Wessex Growth Study). *Arc Dis Child* 1990;**65**:1340–4.
5 **Lipman TH**, Hench K, Logan JD, *et al*. Assessment of growth by primary health care providers. *J Pediatr Health Care* 2000;**14**:166–71.
6 **Waite FL**. Creation of a program to teach health providers accurate measurement of children and an understanding of growth patterns as indicators of health. *Programs and Abstracts of the Pediatric Endocrinology Nursing Society* 1997:23.
7 **Betts PR**, Voss LD, Bailey BJR. Measuring the heights of very young children. *BMJ* 1992;**304**:1351–2.
8 **Johnson T**, Engstrom JL, Haney SL, *et al*. Reliability of three length measurement techniques in term infants. *Pediatr Nurs* 1999;**25**:13–17.
9 **Rosenfeld RG**, Conte F, Arnold M. *Diagnosis and monitoring of growth disorders*. Califon: Gardner-Caldwell SynerMed, 1993.
10 **Laing G**, Rossor EB. Growth screening in schools: an evaluation of the programme in one district. *Child Care Health Dev* 1996;**22**:11–17.
11 **Hearn S**, Majrowski W, Jenkins S, *et al*. Setting up a height surveillance programme: recommendations based on the Hackney growth initiative. *J Med Screen* 1995;**2**:157–9.
12 **Voss LD**, Bailey BJR. Equipping the community to measure children's height: the reliability of portable instruments. *Arch Dis Child* 1994;**70**:469–71.
13 **Cotterill AM**, Majrowski WH, Hearn SJ, *et al*. Assessment of the reliability of school nurse height measurements in an inner-city population (the Hackney growth initiative). *Child Care Health Dev* 1993;**19**:159–65.
14 **Voss LD**, Mulligan J, Wilkin TJ, *et al*. Poor growth in school entrants as an index of organic disease: the Wessex Growth Study. *BMJ* 1992;**305**:1400–2.
15 **Green AA**, MacFarlane JA. Method for the earlier recognition of abnormal stature. *Arch Dis Child* 1983;**58**:535–7.