EVALUATION OF A KNEMOMETER

by

Ekkehard Zöllner MBChB(UCT), DCH(SA), DTM+H(Wits), DPH(Wits)

Submitted in partial fulfilment of the requirements for the

degree of Master of Medicine in Paediatrics

Supervisor:

Professor François Bonnici MBChB(UCT), MMed(UCT), FCP(SA), CDE

Head, Endocrine Unit,

Departments of Paediatrics and Medicine, Groote Schuur Hospital

1996

Contents

Acknowledgements

I would like to express my sincere gratitude to Prof François Bonnici whose interest in growth stimulated my own reading. He was a source of encouragement and constant support over the last months.

I am indebted to Professors David Beatty and Malcolm Bowie for constructive criticisms. Dr Stephan Maritz and his colleagues from the Medical Research Council gave invaluable input. Prof Christopher Vaughan and Mike Price of the Institute of Biomedical Engineering kindly provided the Vernier and supplied precision cut rods.

A word of appreciation to Dr Steve Delport and my other colleagues in Ward G25 who helped in many ways, and to Sheryl Rayner for typing the protocol. Many thanks to Sue Botha for efficient secretarial assistance.

Summary

Knemometry is a sensitive technique to measure lower leg length. It has been utilised in the assessment of short-term growth. Knemometry is compared to other indicators of early growth. Its uses and limitations are highlighted.

The acquisition of a knemometer necessitated a study to establish the reliability and the validity of its measurements, to compare two different recommended measuring procedures and to establish whether repositioning the child between readings is necessary.

Forty-four healthy children ranging in age from 5,2 - 10,9 years were enrolled. Measurements were done by a single observer. The lower leg of each child was measured by a series of 3x3 readings called "measuring procedure 1" (MP1), followed by a series of 4 readings, the child being repositioned between each reading, called "measuring procedure 2" (MP2). Ten metal rods were measured at the start of each measuring day (314,90 - 449,90mm) to assess validity of the measurements.

The mean standard deviation (SD) of the measurements performed on children was 0,4mm and the coefficient of variation (CV) 0,1%, irrespective of the measuring procedure. Rod measurements yielded corresponding values of 0,13mm and 0,03%. The SD tended to increase with increasing length of rods (rod size 404,90 to 449,90mm). The measurement bias was -0,26mm, except for rod size 345mm (+0,07mm). The standard error of the mean when measuring children was lower for MP2 (0,20mm) than for MP1 (0,25mm). Outliers were distributed at random among the readings. Data analysis by one way ANOVA showed that it may be advantageous to reposition the subject in between readings.

The results confirm that the knemometer is a sensitive instrument. Most of the variation in the measurements was attributable to the subjects, rather than the instrument. This could be explained by unidentified minor movements and by the relative "elasticity" of the children compared to the rods. The knemometer is less reliable in the upper range of measurement when rods are measured. The bias is not consistent, reflecting a possible weakness in the construction. Inspection identified

the coupling arm of the knee plate as a possible source. Because outliers cannot be excluded while performing routine measurements, the 95th percentile for SD affords reasonable protection against their inclusion. It is recommended that 4 independent readings should be taken to establish a measurement and that the first reading should not be excluded. The whole series of readings should be repeated if the SD is >0,65mm. Users need to be aware of the construction flaw in the make of the instrument tested.

Introduction

Knemometry was introduced by Valk et al (1983). The knemometer derives its name from the Greek words κνημη, the leg and μετρο, to measure (Hermanussen M 1988).

The lower leg measuring device or knemometer measures the distance from the top of the knee to the sole of the foot with the subject in the seated position. It has the potential to detect lower leg growth over a short period of time because measurements to a hundredth of a millimetre can be performed.

Traditionally linear growth is assessed by taking two height measurements with a stadiometer one year apart (Milner RDG et al 1979, Parkin JM 1989, Hindmarsh PC and Brook CGD 1995). Most investigators agree that the period of assessment should not be shortened in the case of prepubertal children (Marshall WA 1971, Milner RDG et al 1979, Neyzi O et al 1993). Three monthly and six monthly estimations do not correlate well with measurements taken at yearly intervals because of seasonal variation, "the cyclical nature" of prepubertal growth and because of unavoidable amplification of measurement errors when extrapolating short-term growth increments (Butler GE et al 1990, Neyzi O et al 1993, Wales JKH and Gibson AT 1994).

With the advent of recombinant human growth hormone (hGH) in 1985 hGH therapy became more readily available (MacGillivray MH 1995). It now became feasible to treat children with conditions other than growth hormone deficiency eg Turner's syndrome or genetic short stature. On treatment, height velocity in these conditions is improved. Except for patients with Turner's syndrome, the final height does not differ from the pretreatment target height (Drug and Therapeutics Committee of the Lawson Wilkins Pediatric Endocrine Society 1995). Because the administration of hGH is costly, invasive, labour- and time-intensive, prediction of an early growth response becomes imperative. Several early markers or indicators of growth have been investigated. How does knemometry compare to these?

Early Indicators of Growth

Growth hormone deficiency (GHD) was found to be the best predictor for a significant response to hGH albeit with considerable variability (Milner RDG et al 1979). However, when normal short children are investigated the GH response by definition is normal. Other biochemical markers or growth velocity of isolated parts of the skeleton therefore need to be assessed (Wales JKH and Gibson AT 1994).

The acute metabolic tests of nitrogen, urea and calcium excretion were not significantly correlated to height velocity acceleration (Clayton BE et al 1971). Similarly, increased muscle and fat width was also found to be a poor predictor of the annual growth reponse (Milner RDG et al 1979).

Other biochemical markers under investigation are insulin-like growth factors and their binding proteins. Insulin-like growth factor one (IGF-1) and IGF-binding protein three (IGFBP-3) were evaluated, but were found to fluctuate with age and sexual development (Costin G and Kaufman FR 1987, Hasegawa Y et al 1996). Neither basal nor GH-stimulated IGF-1 levels correlate well with subsequent growth in non-GHD patients (Grant JA et al 1984, Geminer JM et al 1984). Similarly, IGF-2 was found to be no better than IGF-1 in its predictive capability.

Mortensen et al (1991) formerly compared the predictive response of IGF-1 with that of knemometry in children with no or partial GHD (Mortensen HB et al 1991). They found that the average lower leg velocity for the first 3 months was significantly correlated with the total body height velocity during the following 9 months ($r=0,75$; $p=0,008$). In contrast, IGF-1 during the first month of treatment was not significantly correlated to height velocity during the first year. Similarly, Hermanussen et al (1989) found IGF-1 of no predictive value, except for one case of complete GHD. Knemometry, however, was found to be predictive of the annual growth velocity in 21 of their 24 short children treated with hGH. Oppizi et al (1989) also found that the pretreatment lower leg velocity increased significantly on hGH therapy, while no significant correlation was found with basal IGF-1 levels.

A low pretreatment IGFBP-3 standard deviation score was significantly correlated ($r=0,8$; $p<0,005$) to improved growth during the first year of hGH therapy in a heterogenous group of patients with GHD and normal short stature (Hasegawa Y et al

1994). Although these results look promising, the clinical utility of IGFBP-3 with respect to non-GHD patients is not yet established. As yet, IGFBP-3 has not been compared to knemometry.

Several bone markers of growth have been investigated. These include procollagen I C terminal propeptide (PICP), alkaline phosphatase (AP), cross-linked telopeptide of type I collagen and osteocalcin (Johansen JS et al 1990, Crofton PM et al 1995). Procollagen type III N-terminal propeptide (PIIINP) as an index of soft tissue growth has also been under investigation (Danne T et al 1989, Crofton PM et al 1995). Only two markers, ie AP and PIIINP were found to contribute significantly to the prediction of an adequate one year height velocity response. However, their combined predictive capability of 59% only marginally improved that of the 3 month statural height velocity response (53%). AP on its own compared poorly to knemometery (Mortensen HB et al 1991).

With the publication of percentiles for roentgengraphic bone lenghts, it became possible to measure bone length directly by serial X-rays (Maresh MM 1955). Aronsen and Selvik refined this technique by employing Roentgen stereophotogrammetry. A surgical approach is utilised. Two tantalum markers are implanted on each side of a growth zone and serial X-rays are taken to assess growth between these markers. The methodological error when assessing rabbit tibial growth is 13 micrometres. Clinical experience however is confined to a few cases investigated in Sweden (Aronsen AS and Selvik G 1988, Hildebrand H et al 1991). Although very sensitive, its clinical use is limited because it is invasive and ionising radiation is utilised in serial measurements.

The arm-span is thought to correlate closely with stature in normal children (Belt-Niedbala B et al 1986). However, standards and growth curves for both arm-spans and arm-lengths are not available. A substantial improvement to conventional methods of measuring the forearm came with the invention of the ulna measuring device, invented by Valk (Valk IM 1971). The reported mean standard deviation (SD) of a series of repeated measurements performed on the ulna is 0,21mm. In spite of its sensitivity, this device never enjoyed widespread acceptance.

Uses and Limitations of Knemometry

Lower leg measurements performed on the knemometer yielded a mean SD of 0,09mm which was substantially lower than the mean SD of serial ulna measurements performed on the ulna measuring device (0,21mm) and comparable to the error of Roentgen stereophotogrammetry. Because knemometry is non-invasive, it is a very attractive tool to use in short-term growth studies.

Wales and Milner determined the growth response to hGH therapy in children with Turner's syndrome and normal variant short stature. They compared the one month response as determined by knemometry, with the yearly height velocity after the children had been treated for 6 months. They found that knemometry has a sensitivity of 93%, a specificity of 55%, a positive predictive value of 72% and a negative predictive value of 86% (Wales and Milner 1990). This implies that the absence of a short-term growth response to hGH, as assessed by knemometry, is a reasonable predictor of long-term treatment failure.

Wolthers and Pedersen have monitored the short-term growth suppression qualities of various forms, routes and doses of glucocorticoids by knemometry. Growth suppression ranging from 0,25 - 0,63mm / week was identified (Wolthers OD and Pedersen S 1990, 1991, 1992, 1993). Short-term knemometry was also found to be more sensitive than 24 hour urine cortisol excretion in detecting systemic effects of exogenous glucocorticoids in children (Wolthers OD and Pedersen S 1995).

Knemometry has also been effectively utilised in other disease states eg Crohns disease (MacKenzie CA and Wales JKH 1993), renal transplant patients (Seidel C et al 1991), adolescent pregnant girls (Michaelsen KF 1994) and anorexia nervosa (Hermanussen M et al 1987). The neonatal knemometer was effective in identifying a decrease in growth velocity in premature babies on dexamethasone therapy (Gibson et al 1993).

The use of knemometry could be extended to include other fields eg assessment of short-term growth response in nutrition intervention or the assessment of growth suppression caused by cytotoxics and immunosuppressives. Within the field of endocrinology the growth response to anabolic steroids or GH releasing hormone as well as the growth suppressive effects of hormonal therapy could also be evaluated.

Knemometry is undoubtedly a useful tool to detect both growth stimulation and suppression in response to various agents. However, lower leg growth was found to be non-linear. Intra-daily fluctuations ranging from -1,5 to 1,5mm (which far exceeded the error of the device) remained apparent even after confounding variables, eg physical exercise or intercurrent illness, were excluded. Furthermore, sudden bursts of lower leg growth called "mini growth spurts", alternating with periods of low growth velocity every 30 - 55 days, were also found (Hermanussen M et al 1988). It became clear that knemometry could not be used in the assessment of the linear growth rate. If an attempt at prediction is made, a 6 month observation period is found to have the highest predictive value (r^2=0,732) for annual height growth (Dean HJ et al 1990). Interpretation of lower leg growth velocity is further complicated by a poor correlation of tibial growth and peak height velocity (Roche AF 1974). Some authors therefore believe that knemometry should never be used to predict the long-term growth response (Wales JKH and Gibson AT 1994). However, knemometry remains the most acceptable technique to monitor short-term growth and detect a response to treatment.

Knemometry in Cape Town

Short stature is the second most common problem seen at the Endocrine and Metabolic Service at Red Cross Children's War Memorial Hospital. Owing to the high cost of hGH, the number of children who can be offered this therapy is limited. Furthermore, due to budgetary constraints patients receive hGH intermittently rather than continuously. Being able to differentiate responders from non-responders has obvious financial implications. As we have seen, a knemometer fulfils this function with reasonable certainty. The use of knemometry, however, is limited by age (MacKenzie CA and Wales JKH 1993, own observations). Below the age of 4 years measurement is technically more difficult, because the child is less co-operative. Wales suggested that "oscillations of growth velocity" appearing at puberty may complicate assessment of real growth during puberty. An analysis of the age spectrum of short children seen at the Endocrine and Metabolic Service revealed that 44% of children are between 4 and 10 years old and could benefit from knemometric assessment.

The Endocrine Service has recently acquired a knemometer which was developed jointly by the FORCE Institute, the Royal Veterinary and Agricultural University and the University of Copenhagen. Although this device differs in design from the one developed by Valk et al, the principles underlying the measurement of lower leg length are the same. The "technical error" (TE) of the Valk knemometer has been reported to be between 0,09 and 0,16mm (Valk IM et al 1983, Hermanussen M et al 1988). The accuracy improved if the first estimation of a series was excluded from the analysis. The TE of the newer Danish knemometer is said to range between 0,3 and 0,4mm (Michaelsen KF et al 1988). Using the Valk knemometer, Hermanussen et al found that 6 independent estimations were optimal for an accurate determination of lower leg length. There was, however, only 0,055mm difference between the median of 6 and 3 estimations. They therefore recommended that 4 readings should be done, of which the first one should be excluded from the analysis, because it was less accurate (Hermanussen M et al 1988). Others, using the same knemometer, have tested for outliers and replaced them when they occurred (Wit JM et al 1987). The manufacturers of the Danish knemometer stipulate that a series of 3x3 readings, with the child being repositioned only after a set of 3 readings, should be taken (Manual for Knemometer). They recommend that the whole series of readings should be repeated if the range or standard deviation is too high. No formal comparison between the measuring procedure recommended by the manufacturers and that recommended by Hermanussen et al has been reported.

Objectives

A study was performed with the following objectives:

1. to establish the reliability of the Danish knemometer by determining the standard deviation (SD), the standard error of the mean (SEM) and the coefficient of variation (CV) of the measurement when measurements were performed by a single observer;
2. to establish the validity of the measurement by the knemometer;
3. to compare the measuring procedure recommended by the manufacturers to the one recommended by Hermanussen et al;
4. to establish whether it is essential to reposition the child in between readings.

Definitions, Subjects and Methods

Definitions

Lower leg length (LLL) is defined as the distance from the sole of the foot to the top of the knee, when the knee is flexed.

A reading is a single determination of LLL.

Lower leg length measurement (LLLM) is the mean of a series of readings.

Standard Error of the Measurement (SEM) = $SD/\sqrt{n}$, where n indicates the number of readings taken from one subject (Prins I 1995).

Coefficient of Variation (CV) = SD/mean (Prins I 1995).

The validity of the measurement by the knemometer is the closeness of these measurements to the "true" LLL (Voss LD et al 1990).

Subjects

Forty-four healthy children (22 boys and 22 girls) attending the youth group of a church, situated in a middle-class suburb in Cape Town were enrolled. Their ages ranged from 5,2 to 10,9 yrs (mean 7,2 yrs), the oldest girl being 9,8 years. The racial composition was as follows: White 36 (82%), Coloured 7 (16%), Chinese 1 (2%).

Methods

Written informed consent from the parents and approval by the University's ethical committee was obtained prior to the commencement of the study.

The measurements were done by a single observer. The children were measured in the afternoon between 12h00 and 17h00, except for 4 children who were measured in the morning between 10h00 and 12h00. No medical history was taken, no clinical examination was performed and the bone age was not estimated.

The measuring procedure was followed as laid down by the manufacturer in the manual. In brief, the child sat on a chair with the right foot resting on a mobile foot plate (see fig 1). The height of the chair was adjusted so that the knee formed an angle approximately 60 to 70 degrees. A Plexiglass knee plate was lowered onto the knee. A Digimatic scale unit was brought into contact with the coupling arm of the knee plate. The foot plate was then manually moved by the investigator, extending the knee. During this movement the maximum length of the lower leg was recorded. A set of 3x3 readings was taken. The child was only repositioned after a subset of 3 readings (referred to as triplet hereafter). Repositioning entailed getting the child up and letting him walk across the room. This whole procedure was called "measuring procedure one" (MP1). "Measuring procedure two" (MP2) consisted of 4 independent readings, with the child being repositioned after each reading.

Fig 1: The Knemometer

The data were recorded on a data capture form (see appendix 1).

The <u>reliability</u> of the measurement was optimised by the following:

1. The child was optimally positioned at each estimation ie the child was placed exactly in the midline against the back of the seat.

2. The scale unit was placed in the bottom position after each LLLM and the display brought to zero.
3. The digital display of the scale unit was covered with a piece of paper after the instrument was zeroed so that the investigator was unaware of each child's baseline reading (which does not correspond to the zero position).
4. If the child moved during the recording of a reading, the particular reading was ignored and a new reading taken.

The validity of the measurements was assessed by measuring control metal rods of different lengths at the start of each measurement day. Starting with the shortest, each metal rod was placed vertically on the foot plate. The knee plate was then lowered on to the top of the metal rod. The Digimatic Scale unit in turn was brought into contact with the coupling arm of the knee plate. A single reading was taken and the findings were recorded on a validity form (see appendix 2). This procedure was repeated with all subsequent rods.

Statistical analysis was performed with the aid of the Minitab computer programme. The data were checked for obvious outliers by inspecting dotplots. A one way ANOVA (analysis of variance) of the three triplets of MP1 was performed to establish whether there was significant variation between the triplets. The SD of the mean of each triplet, SD of the triplet means, the SD of the means of MP2 as well as the SEM and the CV were computed. The validity data were analysed likewise.

In a separate evaluation, measurements of the knemometer were compared to measurements of a Vernier (accurate to a hundredth of a millimeter). The knee plate was lowered on to the coupling arm of the knee plate. Comparative measurements started at the zero position of the Digimatic scale unit. Measurements were done by two observers. The one observer adjusted the Vernier in 15 or 10mm steps while the other one recorded the readings of the knemometer.

A considerable "tremor" in the readings was noted. When different pressures were applied to the Digimatic scale unit, different readings resulted. Examining the scale unit knee plate measuring system revealed a weakness in the coupling arm of the knee plate. Considerable movement of the coupling arm was noticed. If the arm of the Digimatic scale unit was pressed down firmly on to the coupling arm of the knee plate, the slack of the coupling arm was not taken up (see fig 2). Applying manual pressure to the arm of the Digimatic scale unit eased the "tremor", but did not eliminate it. This tendency was more apparent in the lower range of measurement, ie 330 to 360mm from the foot plate. The excessive movement of the coupling arm of the knee plate could thus constitute a potential design and construction flaw.

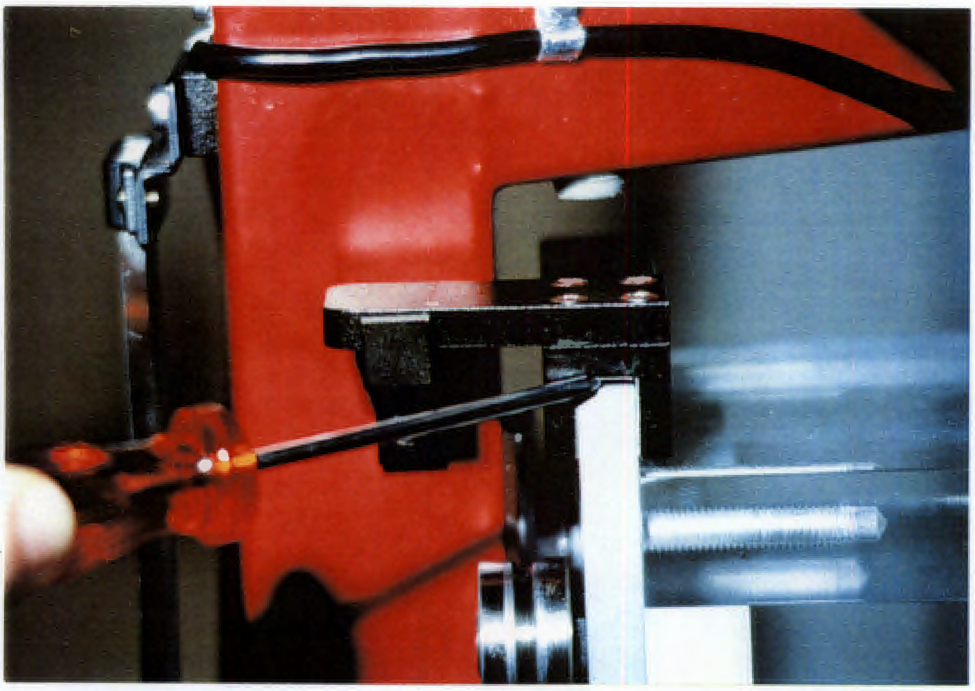

Fig 2: Attachment of the Coupling Arm of the Knee Plate (the skrewdriver points to the slack of the coupling arm)

Results

Inspecting knemometer readings by dotplots revealed 7 outliers - 3 from MP1 and 4 from MP2. The outliers were identified as the first reading in 3 cases only - one from MP1 and two from MP2.

Table 1 illustrates the SD, SEM and the CV with and without outliers being excluded. Note that excluding outliers makes a substantial difference to the result of MP2.

Table 1: Standard Deviations, Standard Errors of the Mean and Coefficient of Variation of different Measuring Procedures

Measuring Procedure	SD (mm)	SEM (mm)	CV (%)
MP1a[1]	0,427	0,247	0,107
1b	0,415	0,240	0,104
MP2a	0,649	0,325	0,162
2b	0,405	0,203	0,101

a = outliers included
b = outliers excluded
1 = SD given as the SD of the triplet means

The SD of the mean of each triplet, the mean SD of the triplets, the SD of the triplet means and the SD of the 4 independent readings are illustrated in Table 2, If there is no significant variation between the triplets, one would expect the SD of the triplet means to be approximately a third of the mean SD of the 3 triplets (because 3 sets of estimations were done). Because the values are quite close to each other, there is real variation between the triplets.

Table 2: Standard Deviations comparing different Measuring Procedures

Measuring Procedure[1]		SD (mm)
MP1	1st triplet	0,375
	2nd triplet	0,383
	3rd triplet	0,443
	mean of triplets	0,400
	triplet means	0,415
MP2		0,405

1 = 7 outliers excluded

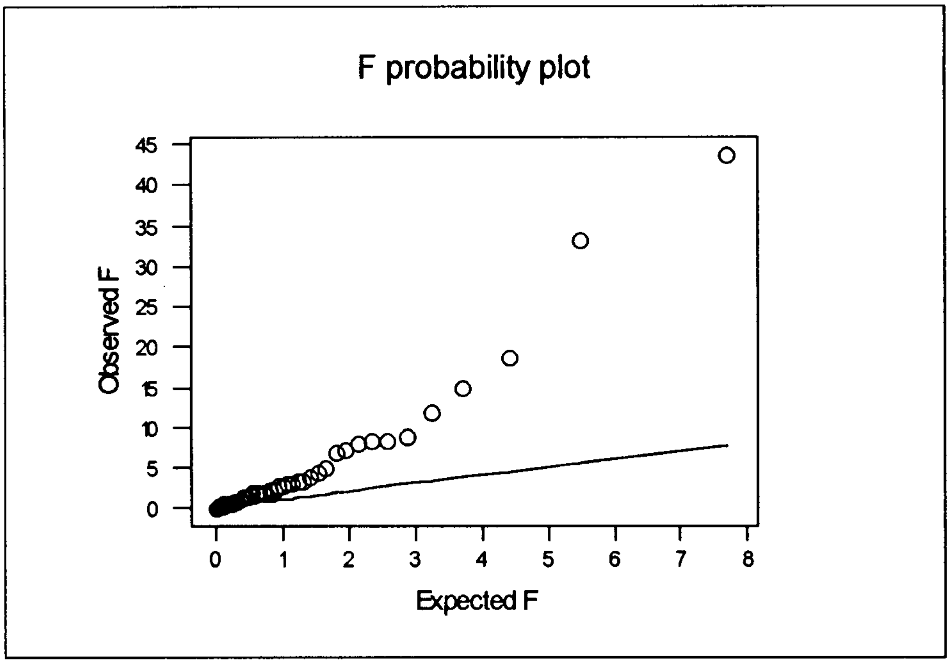

Fig 3: F Probability Plot to Test the Null Hypothesis that there is No Significant Variation between Triplets

To formally test the null hypothesis (Ho) that there is no significant variation between the triplets, a one way ANOVA was performed. This produced 44 observed F statistics whose mean should be 1,5 and whose SD should be 2,60. The observed mean and SD of these statistics are 5,14 and 8,47 respectively, clearly significantly greater than expected ($p<0,0001$). Fig 3 is a probability plot of the F statistics. The plot was obtained by first ranking the 44 observed F values (Rosner B 1986). The order statistics so obtained are the empirical percentiles. Then the corresponding theoretical percentiles of the $F_{2,6}$ distribution were obtained. Plotting empirical versus theoretical percentiles should, under Ho, produce points clustering about a 45^0 line through the origin. Clearly, the observed points depart sharply from expectation. This supports the conclusion based on the mean and the SD of the statistics that Ho should be rejected.

Table 3: Validity and Reliability of Measurements when performed with Metal Rods

Rods (mm)	No of measurements[1]	Mean (mm)	Bias (mm)	SD (mm)	SEM (mm)	CV (%)
314,90	7	315,18	-0,28	0,07	0,03	0,02
329,99	8	330,25	-0,26	0,05	0,02	0,01
345,30	8	345,23	0,07	0,06	0,02	0,02
359,99	7	360,27	-0,28	0,03	0,01	0,01
374,91	6	375,17	-0,26	0,07	0,03	0,02
390,01	8	390,27	-0,26	0,07	0,03	0,02
404,90	8	405,20	-0,30	0,15	0,05	0,04
419,99	8	420,24	-0,25	0,16	0,06	0,04
435,04	8	435,19	-0,15	0,19	0,07	0,04
449,90	7	450,18	-0,28	0,25	0,09	0,06

1 = Outliers excluded

Table 3 shows the measurement bias of the knemometer. The bias was not the same throughout the range of knemometer measurement. Notice that the measurement bias in the third row is out of keeping with all the other results. The mean bias over the whole range of measurement (excluding the bias from row 3) is -0,26mm with a SD of ±0,043mm. This means that the knemometer overreads by 0,26mm, except in the range around 345mm.

The SD tends to increase with increasing length of the rods. This indicates that the knemometer may be less reliable in the higher range of measurements. This phenomenon was not observed in the readings obtained from the children.

The mean SD of all the rod measurements is 0,13mm. Notice that this is approximately a third of the mean SD obtained when measuring children (see Table 1).

Discussion

The results of the evaluation confirm that the Danish knemometer is indeed a sensitive instrument to measure lower leg length. The SD of both MP1 and MP2 are identical to the published TE (Michaelsen KF 1994). But what does this term TE mean?

Many terms have been employed in the literature to describe accuracy of measuring instruments. Valk and others have used statistical terminology, ie the "mean SD" to describe their findings (Valk IM et al 1983, Wit JM et al 1987, Cronk CE et al 1989). Hermanussen reintroduced the term "technical error" (TE) after it had previously been described by L Owen (McCammon RW 1970). Michaelsen defined the TE as "the SD of sequential readings within a measurement" (Michaelsen KF 1994). Hermanussen et al (1988) used this term interchangeably for both the SD and the SEM. The terms "standard error of the measurement", "mean standard error" or "standard deviation of the error" are synonymous to SD in publications by the same authors and others (Hermanussen M et al 1988, Voss LD et al 1990, Dean HJ et al 1990, Seidel C et al 1991). Prins equates the "standard error of the measurement" or the "technical error of the measurement" to the SEM (Prins I 1995). Thus, many terms have been employed to describe the same concepts, while the same term has also been used to describe different concepts.

As is evident, the SD of our knemometer is truly comparable in value and concept to the TE described by Michaelsen. The findings thus confirm Michaelsen's findings that the Danish knemometer is not as accurate as the Valk knemometer or his ulna measuring device. The published accuracy of the Valk knemometer varies from a mean SD as low as 0,09mm (Valk IM et al 1983) to as high as 0,45mm (Dean HJ et al 1990). Interobserver variability may account for some of the difference in the studies that utilised more than one observer.

The SD of the Danish knemometer also compares unfavourably to the "knee height measuring device" (KHMD). The mean SD of 4 readings performed on the KHMD was 0,295mm (Cronk CE et al 1989), suggesting that it is more accurate than the Danish knemometer.

The CV has rarely been quoted in the publications on knemometry. It is indeed not such a useful indicator of reliability, because it is determined by both the SD and the child's mean LLL. The longer the length of the leg, the smaller the CV, provided the SD remains constant. This gives the wrong impression of a higher accuracy in the higher range of readings (Voss LD et al 1990). Our data illustrates an opposite trend, as can be seen from Table 3 (the reason will be discussed later). While measuring children, the CV established with our Danish knemometer is 0,10% which is identical to that found by Wales and Milner who measured children on the Valk knemometer (Wales JKH and Milner RDG 1987). They studied children with an age range of 3,7 to 13,4 years. The mean age of their sample was probably similar to the mean age of our sample, implying that the SD are probably similar and therefore the established accuracies for both knemometers similar.

The SEM, established by utilising six readings on the Valk knemometer was found to be ±0,18mm (Dean HJ et al 1990) or 0,16mm (Hermanussen M et al 1988). The SEM established in our study is 0,20mm which compares very favourably, considering that only four independent readings were taken and not six. This confirms the findings by Hermanussen et al that it is not necessary to perform more than four independent estimations to establish an accurate LLLM.

The SD of the measurements established by measuring rods is much smaller (0,07 - 0,25mm) than the one established by measuring actual children (0,4mm). The reason for this is that children are not inanimate objects and do not tend to sit still for a long time. Even the slightest flicker of movement can make a significant difference to the reading, because the system is so sensitive. Moreover, the lower leg also does not consist only of noncompressable bone, but also of soft tissue, skin, articular surfaces of cartilage and synovial fluid. All these are either compressable or elastic. The state of the soft tissues and the skin will vary with the state of the child's hydration, temperature and activity albeit by a relatively small amount. Consequently, it would be more appropriate to call the mean SD established by measuring rods, the TE. Thus, the TE for the Danish knemometer would be 0,13mm.

In the Wessex Growth study, the reliability of height measurements was assessed (Voss LD et al 1990). It was found that 88-100% of the variance was attributable to the subjects, the rest to the measuring equipment. The subjects' contribution in our study ranged from 83% in the lower range of measurement to 39% in the upper range of measurement (compare SD and CV in Tables 1 and 3). The SD of the rod measurements tended to increase with increasing length of rods. This suggests that the knemometer is less reliable in the upper range of measurements. This tendency was not observed in the patient data. There is no ready explanation for this. A possible explanation might be that bigger children are more likely to sit still than younger ones and consequently contribute less to the overall variance. Alternatively, there may be a construction flaw in the knemometer which becomes more apparent in the upper range of measurement. Possibly, there might be more than one factor accounting for this observation.

The knemometer is also overreading by approximately 0,26mm over most of the measurement range. As the knemometer is used to measure leg length increments over time, a measurement bias that is persistent and of the same magnitude should not have any bearing on the increment. However, when measuring rod size 345,30mm the bias was 0,07mm. This is 0,33mm different from the rest of the measurement range and comes close in magnitude to the established SD. Possible explanations could be transcription error, Digimatic scale unit calibration fault or an inherent structural fault of the scale unit knee plate measuring system.

A potential design and structural flaw was identified in a separate evaluation, comparing measurements of the knemometer to measurements of a Vernier. It is possible that the excessive mobility of the coupling arm of the knee plate could contribute to the discrepancy in the reading around 345mm. The pressure applied by the system could not have been sufficient to take up the slack of the coupling arm. It is suggested that the coupling arm be replaced and the validity re-evaluated.

As is evident from Table 1, outliers have a significant effect on LLLM. However, it is impracticable to test for outliers while performing the measurement. The mini-processor attached to the knemometer is only programmed to exclude the last reading if necessary. A reading that has been recorded and is later identified as an outlier cannot be excluded. As has been shown, the first in a series of readings is not necessarily an outlier. Excluding the first reading does not therefore improve the reliability. Given the constraints of the mini-processor, the only other option is to repeat the whole series of readings if the SD of the readings proves to be too big. What, however, is an acceptable SD and when should the whole series of readings be repeated?

Given that the mean SD of four independent readings is 0,4mm and provided that the observations are normally distributed, then 95% of SDs based on four readings should fall within 0,65mm. This is calculated on the basis of the formula $SD\sqrt{\chi^2_3(0,95)/3}$ where $\chi^2_\nu(\beta)$ is the 100β percentile of a chi-square distribution with ν degrees of freedom. A SD of >0,65mm would thus be unacceptable and the whole series of readings should be repeated. Cronk et al (1989) used the 85th percentile for SD as the tolerance limit. For this study the corresponding value would be 0,53mm. Choosing the 85th percentile rather than the 95th percentile for SD does not necessarily offer more protection against outliers. Chance alone can cause the SD to vary within the 95 percent tolerance limits. Therefore, the upper limit for SD should be 0,65mm. In practice a set of readings should be repeated if the recorded SD of the mini-processor exceeds this value.

The difference between the two measuring procedures is negligible (see Table 1). It would therefore seem immaterial which MP is followed. However, as has been shown, the null hypothesis that there is no significant variation between triplet readings had to be rejected. There is indeed significant variation between triplet readings. This implies that the child needs to be repositioned between readings in order to get a better estimation of the true measurement.

In order to minimise intra-observer variation, the number of readings per LLLM should be as high as possible. As shown, this cannot be achieved by performing readings in triplicate without repositioning the child. It would be better to perform nine completely independent readings, with the child being repositioned in between each reading. However, each child has a limited attention span which in turn limits the number of readings. Hermanussen et al (1988) found that it is difficult to measure a child more than six times. Furthermore, the difference between measuring a child four times or six times was found to be negligible. Generalising from Hermanussen's study, four independent readings is probably all that is required to get an adequate LLLM.

Interobserver variability has not been addressed in this study. This can be minimised if the observers are equally proficient with the technique. No significant variation could be established when adequately trained observers performed measurements with the Valk knemometer (Wales JKH and Milner RDG 1987). The interobserver error of the

KHMD was reported to be 0,33mm which is more than the intraobserver error of 0,20-0,26mm (Cronk CE et al 1989).

Preliminary data assessing interobserver variation of the Danish knemometer showed it to be 0,3mm which is lower than the reported intraobserver variation of 0,4mm (Michaelsen KF et al 1988). This is surprising, as one would have expected the interobserver variation to be higher, because more variables are introduced. This issue needs clarification in a further study.

The use of the Valk knemometer is so complicated that most authorities believe that a considerable period of training is required (Hermanussen M et al 1988, Ahmed SF et al 1995). A learning period of 3 to 4 months has been recommended. Both the Danish knemometer and the KHMD are user-friendly. The investigator of this study was self-taught and in spite of that obtained a SD equal to that published by Michaelsen. This confirms that a relatively short period of training is adequate.

Now that the knemometer has been evaluated, short-term growth of the lower leg can be assessed in clinical trials. Adequate dosing schedules of growth promoting agents can be established. The efficacy of the new GH releasing peptides can be evaluated. Short-term success in nutrition intervention can be documented. The short-term growth suppressive effects of newly released corticosteroids or immunosuppressives to treat atopic, rheumatologic, inflammatory or oncological conditions can be studied.

The knemometer may now also be utilised in clinical practice to aid in the adjustment of doses in tall girls on oestrogen or somatostatin therapy. Its main use in the field of endocrinology will remain the monitoring of the short-term growth response to a growth promoting agent. For example, the knemometer was used to monitor LLL growth velocity of a 10 year old girl with Turner's syndrome treated with hGH and Halotestin. The LLL growth velocity was found to be 0,84mm/week over a period of 3 months. This corresponds to a published range of 0,44-0,93mm/week obtained in girls with Turner's syndrome and treated with hGH (Wales JKH and Milner RDG 1987). Unfortunately no pretreatment data of our patient is available for comparison. These data, however, confirm that the knemometer can indeed identify significant LLL growth in the short-term.

Conclusions and Recommendations

The mean SD and the SEM of 4 independent readings established with the Danish knemometer is 0,4 and 0,2mm respectively. This indicates that the knemometer is a sensitive measuring device with low intraobserver variability. In the upper range of measurement the knemometer is less reliable when measurements are made with metal rods. This tendency is not observed when measuring children.

The knemometer overreads lengths by approximately 0,26mm throughout most of the measurement range. Because the knemometer is used to measure LLL increments over time, a consistent measurement bias will not affect the actual increment. The measurement bias, however, is not consistent which may reflect a weakness in the design and construction of the device.

Outliers have been distributed at random throughout the readings. They have an adverse effect on the SD of the readings. It is not possible to exclude outliers during the measuring procedure.

The measuring procedure of 3x3 readings and the measuring procedure of 4 independent readings yield virtually identical SDs. As the aim in any measuring procedure is not only to keep intraobserver variability to its minimum but also to prevent correlation between readings, the measuring procedure of 4 independent readings is preferred.

The terms "technical error", "standard error of the measurement", "mean standard error", "standard deviation of the error" as used in the literature are not found to be very useful.

The following is recommended:

1. A measuring procedure should be followed where the child is repositioned in between readings.
2. The measuring procedure should encompass a minimum of 4 independent readings and the first reading should not be excluded.
3. The effect of outliers should be minimised by repeating a series of readings if the SD recorded by the mini-processor is >0,65mm.

4. The interobserver variability should be evaluated in a separate study.
5. The construction flaw in the knemometer should be rectified to improve validity and reliability; in the interim users should be aware of this flaw.
6. The confusing terms "technical error", "standard error of the measurement", "mean standard error", "standard deviation of the error" should be abandoned and the purely statistical terms SD and SEM be used instead.
7. The knemometer can now be utilised in trials and clinical practice to establish the growth promotive or suppressive effects of various agents, to monitor LLL growth velocity or adjust doses of different forms of therapy.

References

Ahmed SF, Wallace WHB, Kelnar CJH. Knemometry in childhood: a study to compare the precision of two different techniques. Ann Hum Biol 1995;22(3):247-252.

Aronson AS, Selvik G. X-Ray Stereophotogrammetry, a Method for High Accuracy Analysis of Growth Rate. Acta Paediatr Scand 1988;343(Suppl):186-187.

Belt-Niedbala BJ, Ekvall S, Cook CM, Oppenheimer S, Wessel J. Linear growth measurement: a comparison of single arm-lengths and arm-span. Dev Med Child Neurol 1986;24:343-352.

Butler GE, McKie M, Ratcliffe G. The cyclical nature of prepubertal growth. Ann Hum Biol 1990;17(3):177-198.

Clayton BE, Tanner JM, Vince FP. Diagnostic and Prognostic Value of Short-term Metabolic Response to Human Growth Hormone in Short Stature. Arch Dis Child 1971;46:405-413.

Cole TJ, Freeman JV, Preece MA. Body mass index reference curves for the UK, 1990. Arch Dis Child 1995;73:25-29.

Costin G, Kaufman FR. Growth hormone secretory patterns in children with short stature. J Pediatr 1987;110(3):362-368.

Crofton PM, Stirling HF, Schönau E, Kelnar CJH. Bone alkaline phosphatase and collagen markers as early predictors of height velocity response to growth-promoting treatments in short normal children. Clin Endocrinol 1995;44:385-394.

Cronk CE, Stallings VA, Spender QW, Ross JL, Widdoes HD. Measurement of Short-term Growth with a New Knee Height Measuring Device. Am J Hum Biol 1989;1:421-426.

Danne T, Grüters A, Schuppan D, Quantas N, Enders I, Weber B. Relationship of procollagen type III propeptide-related antigens in serum to somatic growth in healthy children and patients with growth disorders. J Pediatr 1989;114(2):257-260.

Dean HJ, Schentag CT, Winter JSD. Predictive value of short-term growth using knemometry in a large population of healthy children. Acta Paediatr Scand 1990;79:57-63.

The Drug and Therapeutics Committee of the Lawson Wilkins Pediatric Endocrine Society. Guidelines for the use of growth hormone in children with short stature. J Pediatr 1995;127(6):857-867.

Gertner JM, Genel M, Gianfredi SP et al. Prospective clinical trial of human growth hormone in short children without growth hormone deficiency. J Pediatr 1984;104(2):172-176.

Gibson AT, Pearse RG, Wales JKH. Growth retardation after dexamethasone administration: Assessment by knemometer (Abstract). Arch Dis Child 1993;69:505-509.

Grant JA, Howard CP, Daughaday WH. Comparison of growth and somatomedin C responses following growth hormone treatment in children with small-for-date short stature, significant idiopathic short stature and hypopituitarism. Acta Endocrinol 1984;106:168-174.

Hasegawa Y, Hasegawa T, Takada M, Tshuchiya Y. Plasma free insulin-like growth factor I concentration in growth hormone deficiency in children and adolescents. Eur J Endocrinol 1996;134:184-189.

Hermanussen M, Geiger-Benoit K, Sippell WG. "Negative growth" in anorexia nervosa assessed by knemometry. Eur J Pediatr 1987;146:551-564.

Hermanussen M, Geiger-Benoit K, Burmeister J, Sippell WG. Can the knemometer shorten the time for growth rate assessment? Acta Paediatr Scand 1987;337(Suppl):30-36.

Hermanussen M. Knemometry, a new tool for the investigation of growth. Eur J Pediatr 1988;147:350-355.

Hermanussen M, Geiger-Benoit K, Burmeister J, Sippell WG. Knemometry in childhood: Accuracy and standardisation of a new technique of lower leg measurement Ann Hum Biol 1988;15(1):1-16.

Hermanussen M, Geiger-Benoit K, Burmeister J, Sippell WG. Periodical changes of short-term growth velocity ("mini-growth spurts") in human growth. Ann Hum Biol 1988;15(2):103-109.

Hermanussen M, Geiger-Benoit K, Partsch C-S, Burmeister J. Predictive criteria for successful growth promotion in growth hormone therapy of short stature: A comparison between common endocrine parameters and knemometry. Acta Paediatr Scand 1989;78:555-562.

Hildebrand H, Aronson S, Kullendorf C-M, Selvik G. Roentgen Stereophotogrammetric Short-term Analysis of Growth Rate in Children Operated for Crohn's Disease. Acta Paediatr Scand 1991;80:917-923.

Hindmarsh PC, Brook CGD. Short stature and growth hormone deficiency. Clin Endocrinol 1995;43:133-142.

Johansen JC, Pedersen SA, Jørgsen JO et al. Effects of Growth Hormone on Plasma Bone Gla Protein in GH-Deficient Adults. J Clin Endocrinol Metab 1990;70:916-919.

MacGillivray MH. Disorders of Growth and Development. In: Felig P, Baxter JD, Frohman LA, eds. Endocrinology and Metabolism. New York: McGraw-Hill 1995:1621-1625.

MacKenzie CA, Wales JKH. Knemometry and its practical application. Eur Respir Rev 1993;3(13):321-325.

Marsh MM. Linear growth of long bones of extremities from infancy through adolescence. Am J Dis Child 1955;98:725-742.

Marshall WA. Evaluation of Growth Rate in Height over Periods of Less than One Year. Arch Dis Child 1971;46:414-420.

McCammon RW. Human Growth and Development. Springfield: Charles C Thomas, 1970:18-19.

Michaelsen KF, Mortensen HB, Skakkebaek NE. A new knemometer for assessing linear growth (Abstract). Pediatr Res 1988;24:528.

Michaelsen KF. Short-term Measurements of Linear Growth using Knemometry. J Pediatr Endocrinol 1994;7:147-154.

Milner RDG, Russel-Fraser T, Brook CGD et al. Experience with human growth hormone in Great Britain: the report of the MRC Working Party. Clin Endocrinol 1979;11:15-38.

Mortensen HB, Main K, Michaelsen KF, Kastrup KW, Jørgensen JT, Skakkebaek NE. Predicting and maintaining of growth in children with short stature during the first year of growth hormone treatment. Acta Paediatr Scand 1991;80:1150-1157.

Neyzi O, Bundak R, Molzan J, Günöz H, Dasendeliler F, Saka N. Estimation of annual height velocity based on short- versus long-term measurements. Acta Paediatr 1993;82:239-244.

Oppizzi G, Petroncini MM, Marioni T et al. Growth Hormone Treatment of Short Normal Children: a Knemometric Assessment of Effectiveness (Abstract). Acta Paediatr Scan 1989;356(Suppl):150.

Parkin JM. The short child. In: Brook CGD, ed. Clinical Paediatric Endocrinology. 2nd ed. Oxford: Blackwell Scientific Publication, 1989:96-113.

Prins I. Anthropometry in the Newborn: a Clinical Review. Ped Rev Commun 1995;8(3):157-170.

Roche AF. Differential Timing of Maximum Length Increments among Bones within Individuals. Hum Biol 1974;46(1):145-157.

Rosner B. Analysis of Variance. In: Rosner B. Fundamentals of Biostatistics. Boston: PWS Publishers, 1986:442-488.

Seidel C, Schaefer F, Walther U, Schärer K. The application of knemometry in renal disease: preliminary observations. Pediatr Nephrol 1991;5:467-471.

Selvik G, Alberius P, Aronson AS. A roentgen stereophotogrammetric system. Construction, calibration and technical accuracy. Acta Radiol(Diagn) 1983;24:343-352.

Tanner JM, Whitehouse RH, Takaishi M. Standards from Birth to Maturity for Height, Weight, Height Velocity, and Weight Velocity: British Children, 1965. Arch Dis Child 1966;41:454-471.

Valk IM. Accurate measurement of the length of the ulna and its application in growth measurement. Growth 1971;35:297-310.

Valk IM, Langhout Chabloz AME, Smals AGH, Kloppenborg PWC, Cassorla FG, Schutte EAST. Accurate measurements of the lower leg length and the ulnar length and its application in short-term growth measurement. Growth 1983;47:53-66.

Voss LD, Bailey BJR, Cumming K, Wilkin TJ, Betts RP. The reliability of height measurement (The Wessex Growth Study). Arch Dis Child 1990;65:1340-1344.

Wales JKH, Gibson AT. Short term growth: rhythm, chaos, or noise? Arch Dis Child 1994;71:84-89.

Wales JKH, Milner RDG. Knemometry as a Predictor of Response to Somatropin in Turner's syndrome. Acta Paediatr Scand 1987;337(Suppl):37-39.

Wales JKH, Milner RDG. Knemometry in assessment of linear growth. Arch Dis Child 1987;62:166-171.

Wales JKH, Milner RDG. Changes in leg length and height during treatment with Somatropin. Arch Dis Child 1989;64:1708-1712.

Wales JKH, Milner RDG. Predictive Power of Knemometry in Evaluating the Benefit of Somatropin therapy (Abstract). Acta Paediatr Scand 1990;367(Suppl):155.

Wit JM, Van Kalsbeek EJ, Van Wijk-Hoek JM, Leppink GJ. Assessment of the usefulness of weekly knemometric measurement in Growth Studies. Acta Paediatr Scand 1987;76:974-980.

Wolthers OD, Pedersen S. Short-term linear growth in asthmatic children during treatment with prednisolone (Abstract). BMJ 1990;301:145-148.

Wolthers OD, Pedersen S. Growth of asthmatic children during treatment with budesonide. A double blind trial (Abstract). BMJ 1991;303:163-165.

Wolthers OD, Pedersen S. Controlled study of linear growth in asthmatic children during treatment with inhaled glucocorticoids (Abstract). Pediatrics 1992;89:839-842.

Wolthers OD, Pedersen S. Short-term growth during treatment with inhaled fluticasone propionate and beclomethasone dipropionate (Abstract). Arch Dis Child 1993;68:673-676.

Wolthers OD, Pedersen S. Short-term growth in children with allergic rhinitis treated with oral antihistamine, depot and intranasal glucocorticoids. Acta Paediatr Scand 1993;82:635-640.

Wolthers OD, Pedersen S. Measures of systemic activity of inhaled glucocorticoids in children: a comparison of urine cortisol excretion and knemometry. Respir Med 1995;89:347-349.

Appendix 1: Data Capture Form

Data of ....

Date: ....
DOB:..............................
Age: ....
Time:..............................
Sex: M/F
Race: W
C
A
B
Weight. (kg)....................
Height. (cm)....................

Position
Height chair (cm)..............................
Back rest (cm) ....
Foot plate (cm) ....
Knee to foot plate (mm): 300,78

Readings(mm):

MP1 1..
2..
3..
4..
5..
6..
7..
8..
9..

MP2 1..
2..
3..
4..

Appendix 2: Validity Form

VALIDITY

Meas. Day	Rods									
	1	2	3	4	5	6	7	8	9	10