Accuracy and precision of a technique to assess residual limb volume with a measuring-tape

Gustav Jarl

Master’s Thesis, 10 credits, level 61-80 credits
Orthopaedic Engineering
Jönköping, June 2003

Tutor: Magnus Lilja, lecturer
Examinator: Tommy Öberg, professor
Sammanfattning

Volymen på transtibiala stumpar kan förändras dramatiskt efter amputationen och försvåra protesförsörjningen. Skillnaderna mellan olika individer gör det svårt att ge generella rekommendationer om när patienten bör försörjas med definitiv protes. Detta skulle kunna lösas genom att mäta stumpvolymen på varje patient, men de flesta volymuppskattningsmetoder är för komplicerade för kliniskt bruk.

Syftet med studien var att utvärdera validitet och intra- och interpersonella reliabilitet för en metod att beräkna stumpvolymen från omkretsmått på stumpen. I metoden approximeras stumpen som ett antal avklippta koner och stumpänden som en del av ett klot.

Validiteten utvärderades teoretiskt i CAPOD på sex inscannade gipspositiv och manuellt på sex gipspositiv. Reliabiliteten utvärderades genom att jämföra mätningar gjorda av fyra personer på åtta stumpar. Mätningarna gjordes med en trälinjal och ett måttband av metall. Felen uppskattades med intraclass correlation coefficient (ICC), där 0,85 ansågs vara reliabelt, samt det kliniska kriteriet att ett volymfel på ±5% var acceptabelt (5% motsvarar en strumpa).

I teorin visade sig metoden vara valid för alla gipspositiv men i de manuella mätningarna var den bara valid för fyra gipspositiv. ICC var 0,95-1,00 för intrapersonell reliabilitet men bara 0,76 för interpersonell reliabilitet. Både intra- och interpersonell reliabilitet var otillräcklig då kliniska kriterier användes. Felen orsakades av variationer i uppskattningarna av stumpändarnas längder och variationer mellan omkretsmåten.

Metoden behöver utvecklas och är inte lämplig på spetsiga stumpändar. För att förbättra reliabiliteten rekommenderas att använda en lång linjal (ca 30 cm) med en ände som en vinkelhake för att mäta stumpändarnas längder. Att använda ett desinficerbart måttband i kombination med ett fjäderbelastat handtag kan förbättra reliabiliteten på omkretsmåten.

Nyckelord: validitet, mätfel, reliabilitet, proteser, transtibial, volym.
Abstract

Transtibial stump volume can change dramatically postoperatively and jeopardise prosthetic fitting. Differences between individuals make it hard to give general recommendations of when to fit with a definitive prosthesis. Measuring the stump volume on every patient could solve this, but most methods for volume assessments are too complicated for clinical use.

The aim of this study was to evaluate accuracy and intra- and interrater precision of a method to estimate stump volume from circumferential measurements. The method approximates the stump as a number of cut cones and the tip as a sphere segment.

Accuracy was evaluated theoretically on six scanned stump models in CAPOD software and manually on six stump models. Precision was evaluated by comparing measurements made by four CPOs on eight stumps. Measuring devices were a wooden rule and a metal circumference rule. The errors were estimated with intraclass correlation coefficient (ICC), where 0.85 was considered acceptable, and a clinical criterion that a volume error of ±5% was acceptable (5% corresponds to one stocking).

The method was accurate on all models in theory but accurate on only four models in reality. The ICC was 0.95-1.00 for intrarater precision but only 0.76 for interrater precision. Intra- and interrater precision was unsatisfying when using clinical criteria. Variations between estimated tip heights and circumferences were causing the errors.

The method needs to be developed and is not suitable for stumps with narrow ends. Using a longer rule (about 30 cm) with a set square end to assess tip heights is recommended to improve precision. Using a flexible measuring-tape (possible to disinfect) with a spring-loaded handle could improve precision of the circumferential measurements.

Keywords: accuracy, errors of measurement, precision, protheses, transtibial, volume.
Table of contents

INTRODUCTION................................................................................................................1
THEORY..........................................................................................................................4
Measurements..................................................................................................................4
Classification of measurements....................................................................................4
Scales of measurement.................................................................................................5
Errors of measurement...............................................................................................6
The infinite experiment and systematic and random errors..........................................6
Propagation of error......................................................................................................10
Direct measurements....................................................................................................10
Indirect measurements.................................................................................................11
Simultaneous and combined measurements...............................................................13
Determination and correction of errors..........................................................................15
More about accuracy and precision.............................................................................16
Measurements and errors in orthopaedic workshops..................................................18
MATERIAL AND METHODS.........................................................................................21
Study design..................................................................................................................21
Volume determination....................................................................................................21
Determination of volume using circumferential measurements....................................21
Determination of volume using CAPOD® .................................................................22
Accuracy......................................................................................................................22
Theoretical accuracy of tip volume..............................................................................22
Theoretical accuracy of stump volume.........................................................................23
Practical accuracy of stump volume............................................................................23
Precision and sources of error.....................................................................................24
Measuring procedure....................................................................................................24
Intrarater precision........................................................................................................26
Interrater precision........................................................................................................26
Sources of error............................................................................................................27
Subjects.......................................................................................................................28
Statistical methods.......................................................................................................29
RESULTS.......................................................................................................................30
Accuracy.......................................................................................................................30
Theoretical accuracy of tip volume..............................................................................30
Theoretical accuracy of stump volume........................................................................30
Practical accuracy of stump volume............................................................................30
Precision.......................................................................................................................31
Intrarater precision........................................................................................................31
Interrater precision........................................................................................................32
Sources of error............................................................................................................33
DISCUSSION................................................................................................................36
Main results...................................................................................................................36
Accuracy.......................................................................................................................36
Theoretical accuracy of tip volume..............................................................................36
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical accuracy of stump volume</td>
<td>36</td>
</tr>
<tr>
<td>Practical accuracy of stump volume</td>
<td>36</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td><strong>38</strong></td>
</tr>
<tr>
<td>Intrarater precision</td>
<td>38</td>
</tr>
<tr>
<td>Interrater precision</td>
<td>38</td>
</tr>
<tr>
<td>Sources of error</td>
<td>38</td>
</tr>
<tr>
<td><strong>Clinical implications</strong></td>
<td><strong>40</strong></td>
</tr>
<tr>
<td>Methodological considerations and limitations of the study</td>
<td>41</td>
</tr>
<tr>
<td>External accuracy</td>
<td>41</td>
</tr>
<tr>
<td>Internal accuracy</td>
<td>42</td>
</tr>
<tr>
<td>Other methodological considerations</td>
<td>43</td>
</tr>
<tr>
<td>Future studies</td>
<td>44</td>
</tr>
<tr>
<td><strong>CONCLUSIONS</strong></td>
<td><strong>45</strong></td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>46</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>47</td>
</tr>
<tr>
<td><strong>APPENDIX I, INSTRUCTIONS FOR THE MEASUREMENTS</strong></td>
<td>50</td>
</tr>
<tr>
<td><strong>APPENDIX II, MEASURING FORM</strong></td>
<td>51</td>
</tr>
<tr>
<td><strong>APPENDIX III, CALCULATIONS</strong></td>
<td>52</td>
</tr>
<tr>
<td>Error limits for cut cone</td>
<td>52</td>
</tr>
<tr>
<td>Error limits for sphere segment</td>
<td>55</td>
</tr>
</tbody>
</table>
INTRODUCTION

A good fitting of the socket to the residual limb is an important factor to achieve a prosthesis that is functional to the patient (Johansson and Öberg 1998). Both volume and shape are factors influencing the degree of fit between the socket and the stump. Although shape may be a more valid measure of socket fitting, it is much harder to quantify than volume. It is also harder to construct clinical criteria for how stable the shape should be (how big the variations can be and at what sites) before fitting the patient with a definitive prosthesis. The focus of this thesis is the matching of volume between socket and stump.

To get the volume of the socket correct can be especially hard after the amputation when reduction of oedema, atrophy of muscles and weight changes can make the volume of the residual limb change dramatically. These problems make it easier to get the socket fit if one waits with the prosthetic fitting until the volume of the stump has stabilised. On the other hand the patient should be made mobile as soon as possible postoperatively to prevent deconditioning (Bowker et al. 1992). There is also an ambition to start gait training and rehabilitation as soon as possible postoperatively. This situation can be solved by the use of an inexpensive temporary prosthesis during a period postoperatively. (Fernie and Holliday 1982, Lilja and Öberg 1997) The crucial question is when the volume of the stump is stable enough for fitting with definitive prosthesis.

Different authors have investigated the changes in volume after amputation and some have come with recommendations about when the proper time is for fitting with definitive prosthesis (Fernie and Holliday 1982, Persson and Liedberg 1983, Golbranson et al. 1988, Lilja and Öberg 1997).

Fernie and Holliday (1982) measured the volume changes of 18 amputees (17 at transtibial level) with a water displacement technique. The measurement errors of the apparatus were evaluated in a previous study, but only for assessment of cross-sectional areas, not for volumes (Fernie et al. 1978). The results were presented as groups of patients with similar patterns of volume changes. No statistic method was used to separate the patients into different groups. The authors found volume changes not to have any single characteristic pattern and they therefore states that the time for prosthetic fitting cannot be predicted precisely by regular volume measurements postoperatively. The authors recommend fitting with definitive prosthesis after 150 days postoperatively.

Persson and Liedberg (1983) followed 93 transtibial stumps postoperatively. After 12 weeks the mean reduction in volume was 7.3% (standard deviation 10.6%) compared to two weeks postoperatively. The volume of the stump was approximated as a cut cone between circumferential measurements taken proximal and distal at the stump, respectively. The method for assessing volume is questionable but the results indicate big differences between individuals in volume changes postoperatively.

Golbranson et al. (1988) used a water displacement technique to estimate postoperatively volume changes of transtibial amputees. 36 individuals were divided into three groups receiving different treatments to stabilise limb volume. The first group was treated with elastic bandage, the second group had a plaster cast attached to a pylon and a SACH foot (elastic bandage when not ambulating), and the third group had a laminated socket with a
pylon and a SACH foot. The first group showed no significant volume changes over time, but the both latter groups showed significant reductions of volume over time. The correlation coefficients between volume and time were very low, between 0.08 and 0.30 for the different groups, indicating big variations between individuals. Like Fernie and Holliday (1982) the authors found volume stabilisation to be a highly individualistic process. Still, Golbranson et al. (1988) recommend that circumferences are measured postoperatively as a guide for permanent prosthetic fitting. This is almost completely the opposite of the statement by Fernie and Holliday (1982), that the proper time for definitive fitting with prosthesis cannot be predicted by postoperatively volume measurements.

Lilja and Öberg (1997) examined 11 transtibial amputees during 160 days postoperatively. The volume determinations were made with the CAPOD system which measurement errors have been evaluated in an earlier study (Lilja and Öberg 1995). The postoperative treatment consisted of fitting with temporary prosthesis and rehabilitation with physical training, gait re-education, prosthesis training, and occupational therapy. An elastic bandage was used after the surgical dressing had been removed. The authors found all patients to be ready for definitive prosthesis after 120 days, when using a criterion that one stocking is acceptable to wear in the socket. (When the authors used a two stockings criterion, the recommendation was to fit with prosthesis after 100 days) There were big variations between the individuals in the study, the first patient being ready for definitive prosthesis after 80 days and the last after 120 days.

Different postoperative treatments were used in the studies, and the treatments were not always described in detail. Golbranson et al. (1988) evaluated three different methods of stabilising limb volume: an elastic bandage, a plaster cast with a pylon and a SACH foot, and a laminated socket with a pylon and a SACH foot. In the study by Lilja and Öberg (1997) an elastic bandage was used after the surgical dressing had been removed and the postoperative treatment consisted of fitting with temporary prosthesis and physical rehabilitation. Persson and Liedberg (1983) do not report more than that the patients were kept in plaster of Paris during the first two weeks postoperatively and that the stitches were removed at three weeks. Fernie and Holliday (1982) do not describe the postoperative treatment at all. There are different philosophies of the postoperative treatment, concerning the dressing of the stump (rigid, semirigid, elastic, or more recently silicon liner), when to ambulate the patient, etcetera. The ideal would be to evaluate each postoperative method and compare the results to each other. Maybe are the volume changes different for different treatments, and maybe could the individual differences be reduced by some methods. The lack of homogeneity of treatments in the studies above questions the possibility of summing up the results from the different studies. On the other hand, although different treatments were used, the results all had one characteristic in common: the big variations between different individuals.

Both Fernie and Holliday (1982) and Lilja and Öberg (1997) give recommendations of when to fit with definitive prosthesis. The problem with general recommendations is that individual differences make the recommendations more or less accurate when applied on the individual. All the quoted studies show big differences in volume changes between different patients (Fernie and Holliday 1982, Persson and Liedberg 1983, Golbranson et al. 1988, Lilja and Öberg 1997). If these kinds of recommendations are accepted in clinical practice there is a risk that some patients get their definitive prostheses too early, which results in a need for a new socket after a short time. There is also a risk that other patients get their prostheses too late, and therefore get a worse rehabilitation than had been possible.
Another way to address the problem, instead of trying to find the one best time for prosthetic fitting, would be to develop a method to assess limb volume that could be used on every single patient. This would make it possible to make the decision of when to fit with a definitive prosthesis based on the patient’s own individual volume changes. Measurements could be recorded regularly after the amputation and be compared to the patient’s previous volumes and to shrinking patterns from scientific studies. Based on this, the proper time for fitting with definitive prosthesis could be selected. For example, if the volume immediately after the amputation is assumed to be 100%, a criterion could be to fit the patient with a definitive prosthesis when the volume change is less than ±5% in three weeks.

Yet most methods for volume determinations used in scientific settings are by different reasons not suitable for routinely clinical use. Water displacement technique (Fernie and Holliday 1982) can be too complicated to use as a clinical routine. Spiral X-ray computed tomography (Smith et al. 1995) and CAD/CAM systems (Lilja and Öberg 1995, Johansson and Öberg 1998) have also been used, but these systems require expensive equipment.

A proper method for clinical volume assessments would have to be simple to use, inexpensive, not very time consuming, and have an acceptable accuracy and precision. It is also an advantage if the method does not require much effort from the patient while most people amputated in the western world are elderly and can be in a poor physical condition.

A simple method to assess transfemoral residual limb volume with a measuring-tape was described by Krouskop et al. (1979). Circumferences were measured at regular intervals and the stump volume was approximated as a number of cut cones between the measurement sites. The tip of the stump was approximated as a segment of a sphere. The method has never, what it comes to the author’s knowledge, been evaluated on transtibial stumps when it comes to errors of measurement.
THEORY

No matter if volume or some other characteristic is measured, a basic understanding of measurements and errors of measurements is essential in scientific work.

Measurements

A measurement can be defined as a way to procure symbols (often numbers) that represent characteristics of an object, an event, or a condition, where the symbols’ relations to each other are the same as the relations between the objects/events/conditions they are representing (Ackoff 1972).

Classification of measurements

In metrology measurements are traditionally classified into direct, indirect, and combined measurements. The combined measurements have more recently been divided into strictly combined and simultaneous measurements. (Rabinovich 1995)

Direct measurements are made by measuring an object with an instrument and reading the results direct on the instrument. Measuring circumferences of limbs with a measuring-tape is an example of direct measuring. The indirect measurements are based on knowledge of the relations between the quantity in interest and other quantities. The other quantities are measured and the wanted quantity is calculated from the results of the measurements. For example can mass and volume be measured to calculate the density of an object, while the ratio of mass and volume is the definition of density. (Rabinovich 1995) Although direct measurements are very common in orthopaedic workshops indirect measurements are probably never performed, at least not manually. CAD/CAM systems may use indirect measurements when calculating volumes of stumps from the scanned coordinates and the approximated shape.

Strictly combined and simultaneous measurements are closely related. In both cases several quantities are simultaneously measured (usually direct) and the values are put into a equation system that has to be solved. For strictly combined measurements, the measured quantities are of the same type. The equation system the results are put into is based on the relationship between arbitrary objects with the same measurable quantity. Assume that the mass of an object is known. The mass of several different objects can then be found by comparing different combinations of objects, and constructing equations from the results. For the simultaneous measurements, on the other hand, the measured quantities are of different kinds and the equations reflect relationships between the quantities in the nature. (Rabinovich 1995) Assume that the relationship between pressure and temperature of a gas is studied. The coefficient describing the relationship can be found by measuring the pressure by different temperatures and solving the equation system formed by the results. Strictly combined and simultaneously measurements are probably never performed in orthopaedic workshops.

The strictly combined measurements can be viewed as a generalisation of the direct measurements and the simultaneous measurements as a generalisation of the indirect measurements. This means that, when it comes to the measurements physical significance, they can only be classified into direct and indirect measurements. Still, when the processing
of the data after the measuring is in focus, it is practical to distinguish between (a) direct, (b) indirect, and (c) simultaneous and combined measurements. (Rabinovich 1995)

Ackoff (1972) classifies measurements in an alternative way and take example of four types of measurements: numbering, counting, ranking, and measuring in a “restricted” sense. When numbering, symbols (letters, numbers, et cetera) are put on objects or events and the symbols are used as identification of the objects/events in the further processing of the data. No arithmetic operations can be performed with the symbols. When counting a number of positive values are put on elements in a class. For example can money of different denomination be counted: 20+0,5+10 = 30,5. The numbers resulting from the counting can be used in all kinds of arithmetic operations. The elements of interest can also be ranked, which means that the elements are put in a specific order depending on a specific relationship between the elements. Although numbers can be used to describe an element’s rank, no arithmetic operations can be performed with the numbers. Finally, there is also a measurement in a restricted sense. These measurements are made with a constant measuring unit. (Ackoff 1972)

Scales of measurement

Depending on the characteristics of the measuring operations used a specific type of measuring scale can be chosen. When elements are numbered a nominal scale is used and when elements are ranked an ordinal scale is used. When measuring in a restricted sense, either an interval or a ratio scale is used. These four scales are the main types of measuring scales, but other types of scales can be obtained by combining the main scales (Ackoff 1972). All four scales have specific characteristics, which influence both the arithmetic and statistical operation possible to use on the material:

1. A nominal scale categorizes elements or groups of elements in different classes where all members in a class have some specific characteristic in common. It is useful to construct the classes so there is a class for every variation of the characteristic wished to classify and so the classes exclude each other, that is, so every element belongs to one class only. There is no specific order or ranking between the classes. (Ackoff 1972) As an example can patients be classified depending on their diagnoses and transtibial stumps depending on their shape (cylindrical, conical, or bulbous).

2. An ordinal scale ranks the elements in an order, so it is possible to state that a variable value is bigger (better, longer, et cetera) or smaller (worse, shorter, et cetera) than another value. The scale does not quantify the differences between the elements (Körner et al.1998). An example of ordinal scale is the sport medals; gold, silver, and bronze. When medical treatments are evaluated orally an ordinal scale is often (unconsciously) used. The patient is asked if he/she experience his/her condition as good/bad, improved/deteriorated, et cetera. Temperatures of limbs can also be ranked as warm, normal, or cold.

3. On an interval scale the elements are ordered and the distances between the elements are known. Usually the distances are the same between all scale steps, that is, the scale is linear, but the scale can also be logarithmic. There is no “natural zero” on the scale, which limits the mathematical operations possible to use. It is possible to calculate sums and differences, but not products and quotes. It is not possible to state that, for example, 20° Celsius is twice as warm as 10°, while the position of the point zero is arbitrary decided. Still, the most arithmetic operations can be performed with the differences between the values. (Ackoff
1972, Körner et al. 1998) To the author’s knowledge, the interval scale is not used in orthopaedic workshops.

4. On the last scale, the ratio scale, the elements are ranked, the distances between them are known, and there is an absolute zero on the scale. It is possible to calculate sums, differences, quotients, and products (Körner et al. 1998). An example of a ratio scale is the Kelvin scale, where the zero point is the absolute zero for temperature. 20° Kelvin is then really twice as warm as 10°. Volumes, lengths, and circumferences of limbs and ranges of motion of joints are measured on a ratio scale.

Obviously, the latter scales give more information about the measured variables than the former scales. When using the latter scales more types of statistical operations are also possible to use on the material (Ackoff 1972). The focus in this thesis is the measurements in restricted sense, either used on an interval or on a ratio scale.

**Errors of measurement**

Errors are present in almost all kinds of measurements, no matter what kind of measurement that is performed and what kind of scale that is used. Error of measurement can be defined as a deviation of the result from the true value. If \( \mu \) is the true value of the quantity and \( x \) is the result of a measurement of \( \mu \), the absolute measurement error \( \zeta \) is calculated as:

\[
\zeta = x - \mu
\]

Measurement errors can also be expressed in relative form as a fraction of the true value:

\[
\varepsilon = (x - \mu) / \mu
\]

If the errors are very small, they are usually expressed as fractions of the measured value \( x \). (Rabinovich 1995)

Errors of measurement are usually divided into systematic and random errors, each of them having specific characteristics and influences on the result. There are also errors that cannot be classified into these two main categories: the observer may read the result incorrectly, the observer may slip with the pen when notating the result, there may be error due to the rounding off of the numbers, et cetera. These kinds of errors also influence the final result but will not be treated in this thesis.

**The infinite experiment and systematic and random errors**

When a specific measurement is repeated under the same conditions the distribution of the results changes much at first, when the number of measurements is small. When the number of measurements increases the changes in the distribution become smaller. There is an assumption that if the measurement could be repeated an infinitive number of times the distribution would settle down to a specific shape, called the limiting frequency distribution curve or the parent distribution, Figure 1. If the curve is symmetric the mean, mode, and median are the same value and this value is usually chosen to be the “true” value of the experiment \( \mu \). If the curve is asymmetric it is a matter of convention which value is chosen, but most frequently the mean is used. (Boas 1983, Barford 1985)
The true value of the experiment $\mu$ is not necessarily the true value in reality; there may be a systematic error (also referred to as bias). If the systematic error is constant, it is the difference between the mean $\mu$ and the true value, Figure 1. The systematic error is not always constant; it may also vary periodically or according to some mathematical function and must then be evaluated in other ways. If the systematic error is very small or has been corrected for, $\mu$ can be assumed to be the true value. Accuracy (also referred to as validity) is the positive contrast to systematic error, where a high accuracy reflects a small systematic error. Accuracy can be described as a measure of to what extent a method measures what it is intended to measure (Dawson and Trapp 2001). Although the true value should ideally be used as a reference when evaluating accuracy, it is usually unknown. The result of the measurements is then compared to a reference technique known to be accurate, a “gold standard” (Hulley and Cummings 1988).

Still if $\mu$ is close the true value; there is a discrepancy between the individual measurements and $\mu$. This discrepancy is called the random error and reflects the uncertainty of the result. Precision (also referred to as reliability) is the positive contrast to random error, where a high precision reflects a small random error. Repeating the measurements several times and using a statistical method to estimate the spread of the measurements estimates the precision. Usually the root mean square deviation $\sigma$, also called the standard deviation (SD), is used to estimate precision. The coefficient of variation (CV) is a “normalized” standard deviation, which is calculated by dividing the SD by the mean (Hulley and Cummings 1988). The range of the measurements can also be used ($x_{\text{max}}-x_{\text{min}}$) to estimate precision (Råde and Westergren 1995).

Intraclass correlation coefficients (ICCs) can also be used as estimates of precision (Shrout and Fleiss 1979). The ICC can be defined as the correlation between different measurements of the same object and describes the relative homogeneity of measurements within a “class”. There are three main forms of ICCs and they are all calculated with numbers obtained from an analysis of variance (ANOVA). Assume that a number of raters $k$ measure a number of
randomly selected persons $n$. Assume further that the unit for the analysis is the individual measurements, that is, not means of several measurements.

The first form, ICC(1,1), is used when each person/object is measured by a different set of $k$ raters. The raters are randomly selected from a larger population of raters. ICC(1,1) is calculated as:

$$\text{ICC}(1,1) = \frac{(\text{BMS}-\text{WMS})}{\text{BMS} + (k-1)\text{WMS}}$$

where the components are obtained from a one-way ANOVA, Table 1.

The second form, ICC(2,1), is used when each rater makes measurements on each person/object. The raters are randomly selected from a population of raters. ICC(2,1) is calculated as:

$$\text{ICC}(2,1) = \frac{(\text{BMS}-\text{EMS})}{\text{BMS} + (k-1)\text{EMS} + k(\text{RMS}-\text{EMS})/n}$$

where RMS is the mean square raters (variation within persons but between raters) and EMS is the residual mean square = total MS-BMS-RMS. All components are obtained from a two-way ANOVA, Table 1.

The third form, ICC(3,1), is also used when each person/object is measured by each rater, but when the raters in the study are the only raters of interest. The results of the study are then not possible to generalize to other raters than those participating in the study. ICC(3,1) is calculated as:

$$\text{ICC}(3,1) = \frac{(\text{BMS}-\text{EMS})}{\text{BMS} + (k-1)\text{EMS}}$$

where the components in the equation are obtained from a two-way ANOVA, Table 1. (Shrout and Fleiss 1979, Laschinger 1992)

<table>
<thead>
<tr>
<th>Table 1. Sources of variation of one- and two-way ANOVA.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-way ANOVA</strong></td>
</tr>
<tr>
<td>Source of variation</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Between persons</td>
</tr>
<tr>
<td>Within persons</td>
</tr>
</tbody>
</table>

<p>| <strong>Two-way ANOVA</strong>                                        |</p>
<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Mean Squares (MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between persons</td>
<td>BMS</td>
</tr>
<tr>
<td>Within persons</td>
<td>WMS</td>
</tr>
<tr>
<td>-Between raters</td>
<td>RMS</td>
</tr>
<tr>
<td>-Residual</td>
<td>EMS</td>
</tr>
</tbody>
</table>

There are two ways to define agreement in the analysis. According to the first definition, consistency, measurements are in perfect agreement if they are equal or can be additively transformed to equality. According to the other definition, absolute agreement, values are in
perfect agreement only if they are equal. For example are the paired numbers (2,4), (4,6), and (6,8) in perfect agreement using the consistency definition but not when using an absolute agreement definition. Additive transformation is performed by subtracting the mean of each rater from all individual measurements (the three paired numbers will then all take the form (-1,1)). ICC(1,1) and ICC(2,1) are measures of absolute agreement while ICC(3,1) is a measure of consistency. For a perfect agreement between the measurements the ICC will take a value of 1.00. There is no absolute lower limit for the ICC; it depends on the number of raters and is calculated as -1/(k-1). (Laschinger 1992, McGraw and Wong 1996)

The ICCs can also be calculated with the means of more than one rater’s measurements as the unit of analysis, instead of the single measurements. This approach usually makes the ICC higher, but the results cannot be generalized to individual raters, which is usually the purpose of the study. There have also been proposed other versions of ICCs than those presented here, but they will not be treated in this thesis. (McGraw and Wong 1996)

The previous discussion about errors of measurement is based on the concept of the infinite measurement. In reality the number of measurements is finite, and the true mean and standard deviation of the parent distribution can therefore not be precisely estimated. The measurements made can be viewed as a sample from the parent distribution, Figure 2. The mean of the parent distribution \( \mu \) is then estimated by the mean of the measurements in the sample \( X_n \):

\[
x_n = \frac{(x_1 + x_2 + \ldots + x_n)}{n} = \frac{\sum x_i}{n}
\]

where \( n \) is the number of measurements in the sample. The standard deviation of the parent distribution \( \sigma \) is correspondingly estimated by the standard deviation of the measurements in the sample \( s \):

\[
s = \sqrt{\frac{\sum (x_i - X_n)^2}{n-1}}
\]

(Boas 1983, Barford 1985, Bevington and Robinson 1992)

Relative frequency of measurements

![Figure 2. Parent and sample distribution curves.](image-url)
In the same way as an infinitive number of single measurements can be imagined, an
infinitive number of sets of measurements can be imagined, each set giving a mean value \( X \).
The standard deviation of a single measurement \( \sigma \) describes the spread of \( x \) values around \( \mu \) and how close to \( \mu \) a single measurement is likely to be. Correspondingly, the standard
deviation in the mean \( \sigma_m \) (also referred to as the standard error) describes the spread of mean
values \( X \) around \( \mu \). It also shows how close to \( \mu \) \( X \) is likely to be. The exact value of \( \sigma_m \) is not
known, but it can be estimated by the adjusted standard error \( s_m \), using the formula:

\[
s_m = \frac{s}{\sqrt{n}}
\]

Where \( s \) is the standard deviation of the sample and \( n \) is the number of measurements in the
sample. \( s_m \) is thus dependent of both the standard deviation of the single measurements and
the number of measurements made. The narrower the parent distribution \( f(x) \) is and the more
measurements made, the closer to the true value \( X \) is likely to be. The result of a measurement
can then be presented as the mean \( \pm \) the adjusted standard error:

\[
X \pm s_m
\]

When the number of observations increases in a sample, the sample distribution approaches
the parent distribution, Figure 2. The mean \( X \) approaches the true value \( \mu \) and the standard
deviation of the individual measurements \( s \), approaches the population standard deviation \( \sigma \).
The standard deviation in the mean \( \sigma_m \), on the other hand, approaches zero as the number of
measurements increases. (Boas 1983, Barford 1985)

The probability of an individual measurement \( x \), or a mean \( X \), to be at a given distance from \( \mu \)
can be estimated with the help of the normal distribution. If the parent population \( f(x) \) is
normally distributed and the sample is big enough (about 30) the sample distribution is
approximately normally distributed. The probability of a single measurement to be within \( \pm 1 \)
SD of the mean is then 68.3\%, within \( \pm 2 \) SD 95.4\%, and within \( \pm 3 \) SD 99.7\%. (Gellert et al.
1989) The probability of a mean value \( X \) to be within \( \pm 1 \) SD of the true value \( \mu \) is also 68.3\%,
within \( \pm 2 \) SD 95.4\%, and within \( \pm 3 \) SD 99.7\%. The difference is that the latter probabilities
are true for all functions \( g(X) \), no matter what the distribution is of the parent distribution \( f(x) \),
if the sample is big enough. If \( f(x) \) is normally distributed the function \( g(X) \) is also normally
distributed. If \( f(x) \) is not normally distributed, \( g(X) \) is approximately normally distributed if
the number of measurements in the samples is large enough (about 30), according to the
central limit theorem. (Boas 1983, Eason et al. 1992)

**Propagation of error**

The errors of the measurements have different influence on the final result depending on what
operations that are performed on the results of the measurements. The measurements can in
this context be divided into three groups depending on the processing of the data: (a) direct,
(b) indirect, and (c) simultaneous and combined measurements (Rabinovich 1995).

**Direct measurements**

When performing direct measurements no further calculations are made on the material and
the errors will therefore not be different in the final result than in the measurement itself. Still,
if the rater is not aware of the error of the measurements the person may misunderstand the result to be the “true” value.

**Indirect measurements**

Indirect measurements are distinguished by that the measured quantities are not the quantities wanted in the final result. The results from the measurements are used in further calculations and the errors from the measurements will therefore propagate in the calculations. Assume that \( f(x) \) is a function of \( x \) and \( \Delta x \) is the absolute error of \( x \). The first, linear term in Taylor’s series is then usually accurate enough to evaluate the effect of \( \Delta x \) on \( f(x) \), if \( \Delta x \) is small enough or the higher derivates of \( f(x) \) are small enough. The error in \( f(x) \) can then be approximated by the equation for propagation of error:

\[
\Delta f(x) = f'(x) \cdot \Delta x
\]  \[1\]

Where \( f'(x) \) is the first derivative of \( f(x) \) evaluated in the point \( x \), \( f(x) \), Figure 3. The true value of \( x \) is usually unknown, but an approximate value is usually sufficient and can be found experimentally. This equation is not exact, except when \( f(x) \) is a linear function of \( x \):

\[ f(x) = a + bx \]

The higher derivatives in the Taylor’s series, \( f''(x), f'''(x), \) et cetera, are then zero and equation [1] reduces to

\[
\Delta f(x) = b \cdot \Delta x
\]

which is exact for an error \( \Delta x \) of any magnitude, not only for small errors.

Taylor’s series can also be used to estimate errors when a function includes several variables. Assume that \( f(x,y,z) \) is a function of the variables \( x, y, \) and \( z \) and \( \Delta x, \Delta y, \) and \( \Delta z \) are the absolute errors of the variables, respectively. The effect of the errors on \( f(x,y,z) \) can then be approximated by an extension of equation [1]:

\[
\Delta f(x,y,z) = F_x \cdot \Delta x + F_y \cdot \Delta y + F_z \cdot \Delta z
\]  \[2\]

Where \( F_x = \delta f/\delta x, F_y = \delta f/\delta y, \) and \( F_z = \delta f/\delta z. \) \( F_x, F_y, \) and \( F_z \) are evaluated in the point \( (x,y,z) \) or a point close, found experimentally. This equation can correspondingly to equation [1] be used if \( \Delta x, \Delta y, \) and \( \Delta z \) are small enough or the higher derivates \( (F'_x, F'_y, F'_z, F''_x, F''_y, F''_z, \) et cetera) are small enough. The only exception is when the function is linear, the equation is then exact for errors of any magnitude.

The formulas above are based on the assumption that the function can be approximated as a linear function, Figure 3. If the function is strongly non-linear or the errors of the variables are big, or rather when this is the case simultaneously, this approximation is not appropriate. Adding terms of higher order in the Taylor’s series to the equation can solve this. (Arnér 2002)
It is also possible to estimate the standard deviation of a function from the standard deviations of the measured variables. Assuming that the errors in the variables $x$, $y$, and $z$, are uncorrelated, the standard deviation of the function $f(x,y,z)$ can be calculated with the formula:

$$\sigma_{f(x,y,z)} = \sqrt{(\frac{\partial f}{\partial x} \sigma_x)^2 + (\frac{\partial f}{\partial y} \sigma_y)^2 + (\frac{\partial f}{\partial z} \sigma_z)^2}$$  \hspace{1cm} (3)

where $\sigma_x$, $\sigma_y$, and $\sigma_z$ are the standard deviations of $x$, $y$, and $z$, respectively. Equations [2] and [3] can be used on functions with more than three variables, simply by adding more terms to the equations. Equation [2] can be summarized as:

$$\Delta f(x_1,x_2,x_3,...,x_n) = \sum \left( \frac{\partial f}{\partial x_i} \Delta x_i \right)$$

Equation [3] can correspondingly be written in the shorter form:

$$\sigma_f = \sqrt{\sum \left( \frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2}$$

(Bevington and Robinson 1992, Deming 1943, Gellert et al. 1989)

In reality, if the standard deviations of the measured quantities and the formula $f(x)$ are known, the approximation formulas above are not necessary to use. The exact values of $f(x)$ can then be calculated for each measured value and the standard deviation of $f(x)$ can be calculated exactly without using any approximation. Earlier the approximation formulas were useful when the function $f(x)$ was complicated, but after the introduction of computers this is not a problem anymore. Still, there is a use of these approximations when the function $f(x)$ is unknown and is to be estimated by simple linear regression analysis. (Arnér 2002)

It is the belief of the author that indirect measurements are not a well-known area to most certified prosthetists and orthotists (CPOs). To estimate the error of the result demands both that the error of the measurements is known and that the propagation of the error is correctly
calculated. The alternative way, to calculate the errors from the final result, is more exact and less complicated, and is therefore recommended.

**Simultaneous and combined measurements**

When performing simultaneous and combined measurements the results are substituted into an equation system that has to be solved. Usually the number of equations \( n \) is higher than the number of unknown variables \( m \) in the equation system. Because of the errors of measurement the unknown variables cannot be decided so all equations in the system are satisfied simultaneously. The conventional way is then to use the method of least square when choosing a solution of the system. (If \( n = m \), the equation system can only be solved in one way, but there will still be errors in the result) Assume that the equation system has the form:

\[
\begin{align*}
Ax_1 + By_1 + Cz_1 - l_1 &= 0 \\
Ax_2 + By_2 + Cz_2 - l_2 &= 0 \\
&\vdots \\
Ax_n + By_n + Cz_n - l_n &= 0
\end{align*}
\]

Or written more compactly as:

\[
Ax_i + By_i + Cz_i - l_i = 0 \quad [4]
\]

where \( n \) is the number of conditional equations, \( i = 1, \ldots, n \), and \( n > 3 \). \( A, B, \) and \( C \) are the unknown variables wished to estimate and \( x_i, y_i, z_i, \) and \( l_i \) are the results of the \( i \)th series of measurements. When the estimations of the unknowns, \( a, b, \) and \( c, \) are substituted into equation \([4]\) there will be a residual \( v_i \) because the estimations are not exact:

\[
ax_i + by_i + cz_i - l_i = v_i \quad [5]
\]

The values of \( a, b, \) and \( c \) will then be chosen so the sum of the squared residuals gives a minimal value:

\[
Q = \sum v_i^2 = \text{minimum}
\]

If the condition above is to be satisfied it is necessary that

\[
\frac{\partial Q}{\partial a} = \frac{\partial Q}{\partial b} = \frac{\partial Q}{\partial c} = 0
\]

\[
[6]
\]

If equation \([5]\) is derived according to the three derivatives in \([6]\) a system of equations is obtained. Using Gauss’ notation:

\[
\sum x_i^2 = [xx]
\]

the equation system will take the form:

\[
\begin{align*}
[xz]a + [yz]b + [zz]c &= [zl]
\end{align*}
\]

\[
[7]
\]
The solution of a, b, and c, is calculated as:

\[ a = \frac{D_x}{D} \]
\[ b = \frac{D_y}{D} \]
\[ c = \frac{D_z}{D} \]

where the determinant D is:

\[
D = \begin{vmatrix}
[xz] & [zy] & [zz]
\end{vmatrix}
\]

and the determinants \( D_x, D_y, \) and \( D_z \) are found by replacing the first, second, and third columns, respectively, in D with the right hand values of equation system [7]. For example, \( D_x \) will take the form:

\[
D_x = \begin{vmatrix}
[yl] & [yy] & [yz] \\
[zl] & [zy] & [zz]
\end{vmatrix}
\]

(Rabinovich 1995)

All determinants are calculated according to the standard equation:

\[
\begin{vmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}
\]

(Råde and Westergren 1995)

The variance of the conditional equations can be found with the equation:

\[
S^2 = \frac{\sum v_i^2}{n-m}
\]

Where \( n \) is the number of conditional equations in the system and \( m \) is the number of unknowns (in this case three). The variances of the estimated values a, b, and c are calculated as:

\[
S'(a) = \frac{D_{11}}{D} * S^2 \quad S'(b) = \frac{D_{22}}{D} * S^2 \quad S'(c) = \frac{D_{33}}{D} * S^2
\]

where \( D_{11}, D_{22}, \) and \( D_{33} \) are the algebraic complements of the elements \([xx], [yy],\) and \([zz]\) of the determinant D, respectively. They are found by eliminating the row and column in D that the element belongs to. (Rabinovich 1995)

The formulas above are based on the assumption that the conditional equations in the equation system have equal variances, and they cannot be used when the variances are unequal. This can be the case when the different equations in the systems are reflecting measurements made under different conditions, but this will not be treated in this thesis. The method of least
squares is further only possible to use when the conditional equations are linear. When the
equations are non-linear they have to be transformed to a linear form, but this is also beyond
the scope of this thesis. (Rabinovich 1995)

In ortopaedic workshops simultaneous and combined measurements are never performed (if
not inside some CAD/CAM system) and the propagation of these errors are therefore not
relevant to the clinically active CPO.

**Determination and correction of errors**

It is important to be aware of measurement errors, while they are present in all kind of
measurements and influence the results obtained. Errors of measurement are hard to
eliminate, the challenge is rater to reduce the errors within acceptable limits. Systematic errors
can be corrected for, if the magnitude of the error is known. Still, the exact magnitude of the
systematic error is not known and the correction will therefore not be completely correct. This
means that there will always be a small rest of systematic error, even after a correction has
been introduced (Rabinovich 1995).

The systematic error is constant or changes in a regular fashion when the measurement is
repeated. If the error is constant it is the difference between the mean of the measurements
and the true value. This calculation is possible because the random error is symmetrically
distributed (often assumed normally distributed) around an expected value, if the number of
measurements is high, Figure 1 (Huitfeldt 1972). Calculating the mean of the measurements
then eliminates the random error and the only remaining error is the systematic error. The
systematic error may also follow a linear function, i.e. have a constant error in percent. If the
error is constant or linear it can easily be corrected for. If the error follows a more
complicated function it is harder to identify and correct for. A diagram of the results of the
measurements may reveal any regularity in the errors.

The random error is, by definition, random, and the magnitude of the error is therefore not
known on the single measurement. The error can therefore not be corrected for on the single
measurement. If the total random error (measured as the SD) is not constant the situation
becomes even more complex. For example may the random error be different in different
intervals of the scale. These irregularities may be discovered by drawing a diagram of the
results of the measurements. (Huitfeldt 1972) Although the random error cannot be corrected
for on single measurements, improving the measuring procedures can reduce the total random
error. This means that the probability of a measurement to be close to μ increases.

Instead of trying to estimate the errors and introduce corrections after the measurements
(“cure of symptoms”) the measuring process can be analysed and improved so the sources of
errors are eliminated (“cure of causes”). Both systematic and random errors can be reduced in
this way and some improvements reduce both kinds of error in the same time. Both systematic
and random errors are mainly caused by three factors: the observer, the instrument, and the
study subject/object (Hulley and Cummings 1988).

Different observers may operate the instrument in different ways, have a different level of
skill or experience, et cetera. The subjective influence of the observer can be reduced by
mechanize the measuring instrument and by instructing and training the observers on how to
use the instruments. Blinding the observers to different treatment groups and to the results of
other observers is also helpful to reduce errors.
The instrument may be influenced by factors in the environment (heat, moisture, et cetera), the interview technique may be inappropriate for measuring the variable, or a technical instrument may not have been calibrated in a long time.

The studied subjects/objects can also cause errors, for example when there is variability due to biological factors in the subjects (Hulley and Cummings 1988). If volumes of residual limbs are measured, there is a risk that the subjects may contract their muscles during some measurements and therefore introduce an error in the volume. If persons are interviewed they may not always tell the truth or may not remember events correctly.

The environment is sometimes also referred to as a source of error (Ackoff 1972), but is rather an indirect source that has influence on the observer, the instrument, and the studied subject/object. If corrections are to be introduced for the environmental influence the properties of the environment’s influence must be known, for example as a mathematical function. Sometimes this may be well known, as temperature’s influence on expansion of materials, but when the measurement is less technical, as in an interview, the environmental influence may be much harder to quantify and correct for.

No matter what the sources are of the errors, the goal is to keep them small enough so it is possible to distinguish between a difference reflecting an actual change in the variable and a difference depending on errors of measurement.

More about accuracy and precision

The discussion so far has treated accuracy, precision, and errors of measurement from a rather technical point of view. In reality these concepts are quite wide and include much more than have been treated in the previous discussion.

From the general definition of accuracy as a measure of to what extent a process measures what it is intended to measure (qualitatively and quantitatively), many different kinds of accuracy and systematic errors can be identified. In the planning of a medical study there should be concern about the external and internal accuracy, respectively. The external and internal accuracy are schematically illustrated in Figure 4.

The three columns in Figure 4 represent the real world, the study plan, and the performed study, respectively. In the first column there is a wish to get knowledge of a phenomenon in a population in the real world. This leads to the formulation of a research question. By practical reasons the entire target population is not studied. Often a sample is drawn from a population that is easy accessible.

The second column is the theoretical construction of a study to find the answer of the research question. There is an intention to measure some specific variables (reflecting the phenomenon in the real world) on a theoretical sample drawn from the accessible population.

The third column is the study performed, that is, where the study plan is realized. Some measurements are performed on the actual study subjects and results are obtained. Now, if everything works out good, the real study can be performed correspondingly to what was intended in the study plan. Inferences can then be made from the study results obtained on the actual sample to the intended sample (internal accuracy). If the study plan is well constructed
to investigate the phenomenon in the real world and the intended sample reflects the accessible population, the conclusions from the study can be further generalized to this population (external accuracy). If the accessible population further represents the target population in a justifying way, the conclusions can also be generalized to the target population (external accuracy). (Hulley et al. 1988)

The external accuracy is concerned with the questions:
Does the study plan answer the research question?
Is what is stated as “truth” in the study plan possible to generalise to the universe outside the study?
Does the intended sample reflect the accessible population the sample was drawn from?
Does the accessible population reflect the target population?
Are the variables intended to measure good measures of the phenomena of interest?

The internal accuracy is concerned with the questions:
Is the actual study performed in accordance with the study plan?
Do the study results support what is stated as “truth” in the study?
Do the study subjects reflect the intended sample?
Do the measurements performed in the study measure the intended variables (quantitatively and qualitatively), keeping errors of measurement small?

Different forms of accuracy can be put into the main categories external and internal accuracy, respectively. The way to choose the study sample (stratified sampling, cluster sampling, et cetera) is important to get a sample that reflects the accessible population (external accuracy). Content accuracy is a measure of to what extent the items on a test are representing the knowledge wished to investigate (internal accuracy) (Dawson and Trapp 2001).

Other kinds of accuracy may not easily be fitted into the categories external and internal accuracy. Criterion accuracy is a measure of the ability to use the measurement to predict other characteristics related to the measure and cannot easily be defined as either external or
internal accuracy. The construct accuracy is estimated by showing that the used instrument (for example measuring quality of life) is related to other instruments measuring the same characteristic (for example SF-36), and not related to other characteristics (Dawson and Trapp 2001). This is not strictly either external or internal accuracy. The construct accuracy is probably more relevant in survey research than in more “technical” measurements. Still, it is easy to see the parallels to technical measurements where the results are compared to a gold standard, to evaluate accuracy.

Precision can be defined as a measure of the spread or the measurements without reference to the results agreement with the true value. From this definition different kinds of precision and random errors can be identified depending on the cause of the error. Intrarater precision reflects the lack of variability when the same observer repeats the measurements with the same instrument (Dawson and Trapp 2001). This (lack of) variation within the same observer and measuring instrument is by other authors referred to as repeatability (Rabinovich 1995). Interrater precision reflects the lack of variability when different observers repeat the measurement with the same instrument (Dawson and Trapp 2001). This (lack of) variability is also called reproducibility. Reproducibility is also used as a wider concept including variation between raters using different instruments, variation between different laboratories, et cetera (Rabinovich 1995). The different kinds of precision discussed above are possible to estimate for both laboratory and paper scale measurements. For paper scales there are also test-retest and internal consistency precision. Test-retest precision is a measure of the scales ability to reproduce the same measurements on different occasions. By practical reasons this can be hard to administer, and therefore the internal consistency precision is sometimes used as an estimate of the test-retest precision. The internal consistency precision of the items on a scale is a measure of how closely correlated the items are to each other, that is, to what extent the items measure the same characteristic. (Dawson and Trapp 2001)

**Measurements and errors at orthopaedic workshops**

The everyday work at an orthopaedic workshop is highly dependent on qualitative and subjective judgements and the experience of the individual CPO. The experience of examination of patients and production of prostheses and orthoses is hard to communicate to less experienced colleges. This has resulted in a practice at the orthopaedic workshops where the novices can get advice from the more experienced CPOs, but have to learn very much through trial-and-error. The disadvantage of trial-and-error is that it is very time consuming and many patients are bound to get a suboptimal treatment meanwhile.

Standardisation and mechanisation of methods are important to reduce the errors due to the observer, that is, the CPO taking the measures, doing the casting or the rectification, et cetera. The measuring instruments should also be improved to reduce errors. For example may different measuring-tapes give different results. CAD/CAM systems using a laser beam to scan stumps are examples of mechanisation of the casting process, which normally is highly subjective. When the process is less bound to specific persons and instruments, persons and instruments can be exchanged, which makes the process more flexible. For example can another person than the specific CPO does the rectification of the limb model after measurements taken in a standardised manner. Today, measures taken on the patient are so highly influenced by the CPO (and maybe the measuring-tape) that it may be hard for other persons to use them in the production, in case the CPO gets ill, is on holiday, et cetera. Less personal and instrumental influence also makes it possible to localise the whole production somewhere else (central production). A more mechanised procedure is also easier to learn to
perform to the inexperienced CPO. The success of the work is then less dependent on gained experience. The trail-and-error period may then be reduced and fewer patients get an unsatisfactory treatment caused by the inexperience of the CPO. Further, if errors of measurement are evaluated for the standardised methods the material in the CPOs’ journals can be used in scientific studies. Today, many of the measurements performed at orthopaedic workshops have not been evaluated when it comes to errors of measurements, or at least the results are unknown among CPOs.

The crucial question is how the experience and qualitative knowledge of the clinically active CPOs can be transformed into more standardised, mechanised, and quantifiable methods. This development work should preferable be performed by persons with technical knowledge and insight in errors of measurement in cooperation with experienced CPOs.

The focus in this thesis is the quantification of the volume changes after an amputation. Although several factors influence the decision of when to fit with definitive prosthesis, the quantification of volume would at least make the “volume factor” less subjective. Before the method to assess stump volume can be recommended for clinical use it is necessary to evaluate its errors of measurement.
AIM

The aim of the present study was to evaluate accuracy and intra- and interrater precision of a method to assess residual limb volume from circumferential measurements when used on transtibial stumps.
MATERIAL AND METHODS

Study design
The study consisted of two parts, the first part assessing accuracy and the second part assessing intra- and interrater precision of the method to use circumferential measurements for volume determinations.

Volume determination

Determination of volume using circumferential measurements
A technique to assess residual limb volume from circumferential measurements was described and used on transfemoral stumps in an article by Krouskop et al. (1979). The method approximates the stump as a number of cut cones and the stump tip as a tip of a sphere. The only measurements required are circumferences measured at regular intervals and the height of the stump tip. The assumptions in this method are that any two successive cross sections are parallel and circular and that the volume contained between them can be approximated as a right cut cone. The volume between two cross sections can then be approximated by the formula:

\[ V = h^* \frac{C_j^2 + C_k^2 + C_j^*C_k}{12\pi} \]  

[8]

Where \( h \) is the distance between the two cross sections and \( C_j \) and \( C_k \) are the circumferences at the two sections, respectively, Figure 5.

![Figure 5. Approximation of stump volume as a number of cut cones.](image)

Further is the volume of the stump tip approximated as a segment of a sphere. The volume can then be estimated by the formula:

\[ V_{\text{tip}} = \frac{t^3*\pi}{6} + \frac{t*C^2}{8\pi} \]  

[9]
Were \( t \) is the height of the segment and \( C \) is the circumference at its base, Figure 6. The total volume of the stump is calculated by adding all the incremental volumes and the volume of the tip. In the present study the volumes were calculated with either a programmable calculator\(^a\) (Casio fx-7700GE) or a simple program written in Microsoft Excel.

![Figure 6. Approximation of stump tip volume as a sphere segment.](image)

**Determination of volume using CAPOD\(^b\)**

The CAPOD\(^b\) system is a CAD/CAM system developed for prosthetics and orthotics. It consists of a laser scanner unit, CAD software for design rectification of the scanned objects, and a milling machine. The CAPOD system has been described in detail in an article by Öberg et al. (1989).

Volume calculations were performed by taking values of circumferences and distances in the software and substitute them into the formulas [8] and [9]. The software also gives information about the volume contained within any two limits on the scanned stump. This volume was used as a reference, a gold standard, to compare the calculated volumes to.

**Accuracy**

The first part of the study, assessing accuracy of the method, was divided into three steps considered with theoretical accuracy of tip volume, theoretical accuracy of stump volume, and practical accuracy of stump volume, respectively.

**Theoretical accuracy of tip volume**

In the first step the aim was to choose the best site for taking the last circumferential measurement, that is, to choose the height of the tip, \( t \), Figure 6. Six residual limb models were scanned with CAPOD and circumferential measurements were made inside the software, Figure 7. Comparisons were made between tip volumes calculated by the software calculated and tip volumes calculated from circumferential measurements. Comparisons were made between taking the last measurement at five respectively four cm from the distal end of the stump.

\(^a\)Casio Computer Co., Ltd., Tokyo, Japan.

\(^b\)Össur Engineering, Box 67, SE-751 03 Uppsala, Sweden
Theoretical accuracy of stump volume
In the second step, measurements made in the software were used to calculate the whole residual limbs. These volumes were compared to the volumes calculated by the software. As a start the circumferences were measured five cm apart, that is, $h = 5$ cm. If not an acceptable accuracy could be obtained the procedure was repeated with $h = 4$ cm and with $h = 3$ cm, respectively. The limit for an acceptable accuracy was set to $\pm 5\%$ of the reference volume, that is, the volume calculated by the software. 5% is approximately the volume of one stocking (Lilja and Öberg 1997).

Practical accuracy of stump volume
In the third step the second step was repeated in reality on the residual limb models, Figure 7. Every model was measured manually five times by the author with $h = 5$ cm. The procedure was then repeated with $h = 4$ cm and $h = 3$ cm, respectively, if an acceptable accuracy could not be obtained with $h = 5$ cm. The same criterion for an acceptable accuracy was used as in the second step. The volumes of the models had previously been assessed in studies using a water immersion technique. These volume determinations were used as reference volumes to compare the calculated volumes to. The volumes of models 1-3 had been determined by Johansson and Öberg (1998) and the volumes of models 4-6 had been determined by Lilja and Öberg (1995).

Figure 7. Residual limb models 1-6. Anterior view.

An Otto Bock wooden rule and a metal circumference rule$^c$ were used for the measurements, Figure 8. The wooden rule was 20 cm long and had a resolution of 1 mm. The circumference rule was 16 mm wide and could show circumferences with a resolution of 0,1 mm, but in this study a scale with a 1 mm resolution was used and all measurements were rounded off to the closest mm. The metal circumference rule was used instead of the flexible measuring-tape

$^c$Luna AB, Sandbergsv. 3, SE-441 80 Alingsås, Sweden
advantage if the stump has unhealed wounds. Another advantage is that the circumference rule has a construction that makes it easier to measure the circumferences in one plane without twisting the rule.

![Image of circumference rule and wooden rule.](image)

*Figure 8. Metal circumference rule and wooden rule.*

**Precision and sources of error**

The second part of the study, assessing precision, was divided into two steps. In the first step the method was used on transtibial residual limbs and intra- and interrater precision were calculated. In the second step the results were analysed to find the sources of the errors.

**Measuring procedure**

Four CPOs measured eight transtibial stumps in a randomised order, Figure 9. The same kind of equipment was used as when assessing accuracy, Figure 8. The rating CPO measured every stump five times before measuring next stump. Before the measuring began a tape with a pen mark was put on the stump by the author seven cm distal from the anterior, proximal end of tibia, Figure 10. A thin nylon stocking was then put on the stump to protect the skin from the wear and tear of the repeated changes of tapes.
Every CPO was given written instructions for the measurements, telling them to ask the patient to sit with their knees extended and muscles relaxed during the measuring, Appendix I. The CPOs were also given forms to fill in with the results, Appendix II. The CPO put an adhesive tape along the tibial crest and made a mark four cm from the distal end of the stump, Figure 10. Marks were then made every five cm going distal from the anterior, proximal end of tibia. When less than five cm was left to the first mark, the distance was recorded and filled in the form. Circumferential measures were taken by every mark and also filled in the form. A measurement was also made by the mark seven cm distal from the proximal end of tibia. Last, the tape at tibia was removed. This whole procedure was repeated for every measurement. In this way the proximal tibia and the measuring sites had to be identified for every measurement. The tape seven cm distal from proximal tibia was never removed, this mark was “constant” during the whole measuring session. Comparisons could then be made between the precisions of the circumferences when the site had to be identified by palpation and when a given site was used.

*Figure 9. Residual limbs A, G, B, C, D, E, F, and H.*
Intrarater precision

The intrarater precision was estimated in two ways: by calculating the ICC and by calculating the number of measurements within acceptable error limits. The limit for an acceptable ICC was set to 0.85. An acceptable error was considered to be ±5% of the true volume. The clinical criteria for an acceptable intrarater precision were: 1) at least 95% of the measurements within the ±5% limit and 2) no more than one unacceptable volume estimation per stump. The intrarater precision was calculated for every CPO individually.

The ±5% limit used for the evaluation of precision was based on the assumption that the systematic error was zero, or could be fully corrected for. The total error, which should not exceed ±5%, would then only consist of the random error.

As the true volumes of the stumps were not known the mean of the five volumes calculated by each CPO was calculated and used as a reference volume for that CPO on that stump. While the systematic error was assumed to be zero, the mean of repeated measurements would be the true volume. The means and error limits were calculated in the same manner for all stumps and repeated for all CPOs.

Interrater precision

The interrater precision was correspondingly estimated in two ways: by calculating the ICC and by calculating the number of measurements within acceptable error limits. Again the limit for an acceptable ICC was set to 0.85. The clinical criterion for an acceptable interrater precision was that all volumes assessed by the CPOs should be within the error limits, that is, ±5% of the true volume. While the true stump volume was not known the mean of all 20 measurements made on each stump (four CPOs measured every stump five times each) was assumed to be the true volume.

Each CPO’s fifth measurement on each stump was used in the analysis. The fifth measurement was chosen to avoid errors due to introducing problems with the measuring
procedure. Further were means of more than one measurement used in the analysis instead of the fifth measurement alone, to investigate if this approach could improve the precision.

**Sources of error**

In the second step of the precision estimation an attempt was made to identify the sources of the errors and improve the method for future use. It was assumed that errors were mainly caused by two factors: 1) problems with axial measurements, that is, problems palpating the anterior, proximal end of tibia and assessing the height of the stump tip, and 2) problems with circumferential measurements, that is, applying the same tension to the circumference rule every time and holding the rule horizontally.

The first factor, problems with axial measurements, gives rise to variance between the sites where the measurements are taken and variance in stump lengths and tip heights. This was analysed by comparing the estimated stump lengths. Big variations would indicate axial measurements to be a source of error. The first factor was also analysed by comparing the circumferences at the constant mark at the tape never removed to the circumferences measured at proximal tibia and at four cm from the tip end, respectively. The assumption was that a larger variation found between the circumferences at proximal tibia or four cm from the stump end (compared to the circumferences at the constant mark) would indicate problems with correctly identifying these measuring sites.

The second factor gives rise to variance between the circumferential measurements although taken on the same site. This was analysed by comparing circumferences measured at the same site, that is, at the constant mark at the tape never removed.

Limits for acceptable errors of the axial and the circumferential measurements where calculated for the cut cone approximation and the sphere segment approximation, respectively, Appendix III. An axial or a circumferential error causing a volume error of more than ±5% alone was considered too big. The calculations were done with the equation [2], the equation for propagation of error. The most conservative limits were used, that is, the limits allowing the smallest error.

Other problems with the method were also analysed. The approximation of the stump tip volume is not possible to use on tips where the sphere used in the approximation will have a radii that is smaller than the tip height t, Figure 11. The circumference of a circle is calculated as:

$$C = 2\pi r$$

Where r is the radii of the circle. The smallest possible circumference to be used is found by substituting r with t:

$$C = 2\pi t$$

When t = 4.0 cm the smallest allowed circumference is:

$$C = 2\pi*4 = 8\pi = 25.1327\ldots$$

which is rounded up to 25.2 cm. If t = 5.0 cm the smallest allowed circumference is:
\[ C = 2\pi \times 5 = 10\pi = 31.4159\ldots \]

which is rounded up to 31.5 cm.

Subjects
Seven patients, five men and two women, with eight transtibial stumps were included in the study. Letters were sent out to amputees in mainly the Jönköping city area, and those responding were included in the study. Inclusion criteria were 1) transtibial amputation, and 2) at least four months since amputation. Patient data is listed in Table 2. Informed written consent was obtained from all patients participating in the study. Approval to perform the study was obtained from Jönköping University.

The persons performing the measurements were all clinically active CPOs with varying clinical experience. CPO 1 had four months, CPO 2 two years and four months, CPO 3 11 years, and CPO 4 two years and four months of clinical experience.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Sex</th>
<th>Age</th>
<th>Amputated limb (L/R)</th>
<th>Stump</th>
<th>Time since amputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>80</td>
<td>L</td>
<td>A</td>
<td>12 yr</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>83</td>
<td>L</td>
<td>C</td>
<td>2 yr</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>64</td>
<td>L</td>
<td>D</td>
<td>57 yr</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>86</td>
<td>L</td>
<td>E</td>
<td>5 yr</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>73</td>
<td>R</td>
<td>F</td>
<td>3 yr 6 mth</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>71</td>
<td>L</td>
<td>B</td>
<td>5 mth</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>G</td>
<td>7 mth</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>48</td>
<td>L</td>
<td>H</td>
<td>10 yr</td>
</tr>
<tr>
<td>Mean</td>
<td>---</td>
<td>72.1</td>
<td>---</td>
<td>---</td>
<td>11 yr 4 mth</td>
</tr>
<tr>
<td>SD</td>
<td>---</td>
<td>13.1</td>
<td>---</td>
<td>---</td>
<td>18 yr 11 mth</td>
</tr>
<tr>
<td>Range</td>
<td>---</td>
<td>48-86</td>
<td>---</td>
<td>---</td>
<td>5 mth-57 yr</td>
</tr>
</tbody>
</table>

Figure 11. Approximation spheres for different circumferences.
Statistical methods
SPSS (version 10) and NCSS 2000 were used for the statistical analysis. Means, standard deviations (SD), Student’s t-tests, and coefficients of variation (CV) were calculated according to standard procedures. The t-tests were calculated manually. The ICC (form 2,1) was used as estimate of intra- and interrater precision (Shrout and Fleiss 1979) and was calculated with SPSS.
RESULTS

Accuracy

Theoretical accuracy of tip volume

The relative error was smaller on all models when the circumference was measured four cm (t = 4 cm) compared to five cm (t = 5 cm) from the distal end, Table 3. Therefore, four cm from the end was used as the site for the measurement in the further study.

Table 3. Tip volumes of residual limb models (ml or cm$^3$) calculated from measurements in CAPOD and determined by CAPOD itself. Differences in absolute and relative errors.

<table>
<thead>
<tr>
<th>Method</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t = 5 cm</td>
<td>202,45</td>
<td>174,05</td>
<td>256,37</td>
<td>224,22</td>
<td>163,41</td>
<td>215,25</td>
</tr>
<tr>
<td>t = 4 cm</td>
<td>136,48</td>
<td>117,24</td>
<td>178,26</td>
<td>155,59</td>
<td>107,75</td>
<td>146,68</td>
</tr>
<tr>
<td>CAPOD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t = 5 cm</td>
<td>182</td>
<td>156</td>
<td>250</td>
<td>216</td>
<td>142</td>
<td>201</td>
</tr>
<tr>
<td>t = 4 cm</td>
<td>131</td>
<td>114</td>
<td>177</td>
<td>157</td>
<td>102</td>
<td>144</td>
</tr>
<tr>
<td>Calculated – CAPOD (t = 5 cm)</td>
<td>20,45 (11,24%)</td>
<td>18,05 (11,57%)</td>
<td>6,37 (2,55%)</td>
<td>8,22 (3,81%)</td>
<td>21,41 (15,08%)</td>
<td>14,25 (7,09%)</td>
</tr>
<tr>
<td>Calculated – CAPOD (t = 4 cm)</td>
<td>5,48 (4,18%)</td>
<td>3,24 (2,84%)</td>
<td>1,26 (0,71%)</td>
<td>-1,41 (-0,90%)</td>
<td>5,75 (5,64%)</td>
<td>2,68 (1,86%)</td>
</tr>
</tbody>
</table>

Theoretical accuracy of stump volume

When the whole volume of the models was estimated in CAPOD the method showed to be accurate on all models when the circumferential measurements were taken five cm apart, Table 4. It was therefore never necessary to take the measurements any closer to each other.

Table 4. Volumes of residual limb models (ml or cm$^3$) calculated from measurements in CAPOD and determined by CAPOD itself. Differences in absolute and relative errors.

<table>
<thead>
<tr>
<th>Method</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculated h = 5 cm</td>
<td>651,95</td>
<td>690,14</td>
<td>1251,86</td>
<td>937,17</td>
<td>612,72</td>
<td>1395,78</td>
</tr>
<tr>
<td>CAPOD</td>
<td>635</td>
<td>686</td>
<td>1234</td>
<td>921</td>
<td>587</td>
<td>1375</td>
</tr>
<tr>
<td>Calculated (h = 5 cm) – CAPOD</td>
<td>16,95 (2,67%)</td>
<td>4,14 (0,60%)</td>
<td>17,86 (1,45%)</td>
<td>16,17 (1,76%)</td>
<td>25,72 (4,38%)</td>
<td>20,78 (1,51%)</td>
</tr>
</tbody>
</table>

Practical accuracy of stump volume

When comparing the volumes from manual measurements to the volumes from the literature, the results showed that for four of the models five cm between the measurements was enough to reach an acceptable accuracy, Table 5. For models 4 and 5, neither five nor four cm between the measurements gave a satisfying accuracy. For these models the measuring procedure was repeated with three cm between the circumferences, but still the relative error was greater than ±5%. All differences between the volumes but two were statistical
significant. In two cases, model 2 (h = 4 cm), and model 5 (h = 3 cm), one volume estimation was excluded from the analysis when it was obvious that one circumferential measurement was missing in the form.

Table 5. Volumes of residual limb models (ml or cm$^3$) determined by circumferential measurements and from previously studies using water immersion technique. Differences in absolute and relative errors.

<table>
<thead>
<tr>
<th>Method</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculated</td>
<td>mean</td>
<td>803.40</td>
<td>901.80</td>
<td>1280.35</td>
<td>1264.96</td>
<td>1147.55</td>
</tr>
<tr>
<td>(h = 5 cm)</td>
<td>CV</td>
<td>2.03%</td>
<td>0.56%</td>
<td>0.60%</td>
<td>0.28%</td>
<td>0.29%</td>
</tr>
<tr>
<td>Calculated</td>
<td>mean</td>
<td>787.89</td>
<td>882.74</td>
<td>1276.32</td>
<td>1268.44</td>
<td>1141.54</td>
</tr>
<tr>
<td>(h = 4 cm)</td>
<td>CV</td>
<td>0.43%</td>
<td>0.61%</td>
<td>0.48%</td>
<td>0.43%</td>
<td>0.79%</td>
</tr>
<tr>
<td>Calculated</td>
<td>mean</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>1287.56</td>
<td>1142.55</td>
</tr>
<tr>
<td>(h = 3 cm)</td>
<td>CV</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>1.39%</td>
<td>0.37%</td>
</tr>
<tr>
<td>Water immersion</td>
<td>mean</td>
<td>799.38</td>
<td>901.56</td>
<td>1304.41</td>
<td>1202.2</td>
<td>1020.0</td>
</tr>
<tr>
<td></td>
<td>CV</td>
<td>0.66%</td>
<td>0.48%</td>
<td>0.26%</td>
<td>0.53%</td>
<td>0.66%</td>
</tr>
<tr>
<td>Calculated (h = 5 cm) – Water immersion</td>
<td>4.02</td>
<td>0.24</td>
<td>-24.06**</td>
<td>62.76***</td>
<td>127.55***</td>
<td>46.16**</td>
</tr>
<tr>
<td></td>
<td>(0.50%)</td>
<td>(0.027%)</td>
<td>(-1.84%)</td>
<td>(5.22%)</td>
<td>(12.50%)</td>
<td>(3.16%)</td>
</tr>
<tr>
<td>Calculated (h = 4 cm) – Water immersion</td>
<td>-11.49***</td>
<td>-18.82**</td>
<td>-28.09***</td>
<td>66.24***</td>
<td>121.54***</td>
<td>38.95**</td>
</tr>
<tr>
<td></td>
<td>(-1.44%)</td>
<td>(-2.09%)</td>
<td>(-2.15%)</td>
<td>(5.51%)</td>
<td>(11.92%)</td>
<td>(2.67%)</td>
</tr>
<tr>
<td>Calculated (h = 3 cm) – Water immersion</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>85.36***</td>
<td>122.55***</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(7.10%)</td>
<td>(12.01%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = p<0.05  ** = p<0.01  *** = p<0.001. Relative errors larger than ±5% written in fat.

**Precision**

**Intrarater precision**

The ICC ranged from 0.95 to 1.00 for intrarater precision, Table 6. The lower limit of the confidence interval was higher than 0.85 for all CPOs. The clinical criteria for an acceptable intrarater precision were fulfilled by one of the four CPOs, Table 7. The other persons had less than 95% of the measurements within the error limits and more than one imprecise measurement on some stumps.
Table 6. Intrarater precision of stump volumes, ICC.

<table>
<thead>
<tr>
<th></th>
<th>ICC</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower limit</td>
<td>Upper limit</td>
</tr>
<tr>
<td>CPO 1</td>
<td>1,00</td>
<td>0,99</td>
</tr>
<tr>
<td>CPO 2</td>
<td>0,95</td>
<td>0,88</td>
</tr>
<tr>
<td>CPO 3</td>
<td>0,96</td>
<td>0,91</td>
</tr>
<tr>
<td>CPO 4</td>
<td>0,96</td>
<td>0,89</td>
</tr>
<tr>
<td>Average</td>
<td>0,97</td>
<td>0,92</td>
</tr>
</tbody>
</table>

Table 7. Intrarater precision of stump volumes, clinical criteria. Ranges rounded up to closest 0,1%.

<table>
<thead>
<tr>
<th>Stump</th>
<th>Range (% of mean)</th>
<th>Precise measures (within ±5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-1,7 – 3,7</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-11,3 – 12,5</td>
<td>1/5</td>
</tr>
<tr>
<td></td>
<td>-6,3 – 5,5</td>
<td>3/5</td>
</tr>
<tr>
<td></td>
<td>-2,3 – 6,7</td>
<td>4/5</td>
</tr>
<tr>
<td>B</td>
<td>-1,6 – 0,8</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-4,2 – 3,2</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-5,8 – 6,3</td>
<td>2/5</td>
</tr>
<tr>
<td></td>
<td>-5,1 – 5,1</td>
<td>3/5</td>
</tr>
<tr>
<td>C</td>
<td>-2,0 – 2,3</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-2,6 – 2,2</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-1,6 – 1,0</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-2,3 – 1,7</td>
<td>5/5</td>
</tr>
<tr>
<td>D</td>
<td>-1,6 – 1,9</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-5,5 – 5,6</td>
<td>3/5</td>
</tr>
<tr>
<td></td>
<td>-1,9 – 1,1</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-2,5 – 3,7</td>
<td>5/5</td>
</tr>
<tr>
<td>E</td>
<td>-3,0 – 1,7</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-4,5 – 7,3</td>
<td>4/5</td>
</tr>
<tr>
<td></td>
<td>-2,8 – 1,1</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-6,9 – 4,9</td>
<td>4/5</td>
</tr>
<tr>
<td>F</td>
<td>-1,2 – 1,0</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-3,7 – 3,4</td>
<td>4/5</td>
</tr>
<tr>
<td></td>
<td>-2,6 – 4,7</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-4,0 – 3,1</td>
<td>5/5</td>
</tr>
<tr>
<td>G</td>
<td>-1,4 – 1,5</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-3,0 – 1,6</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-1,2 – 1,6</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-4,4 – 7,2</td>
<td>5/5</td>
</tr>
<tr>
<td>H</td>
<td>-2,6 – 2,4</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-2,0 – 1,5</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-4,1 – 2,7</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>-7,1 – 4,2</td>
<td>4/5</td>
</tr>
<tr>
<td>Total precise measures</td>
<td>40/40</td>
<td>33/40</td>
</tr>
<tr>
<td></td>
<td>(100%)</td>
<td>(82.5%)</td>
</tr>
</tbody>
</table>

Errors larger than ±5%, more than one imprecise measure per stump, and less than 95% precise measures written in fat.

Interrater precision

The interrater precision was lower than the intrarater precision and did not fulfil the criterion for an acceptable ICC, that is, 0,85. ICC was 0,76 with a lower limit of the confidence interval at 0,47 when the fifth measurement of each CPO was analysed, Table 8. Using mean values of more then one measurement increased the ICC to as most 0,79. When using the clinical criteria three of the eight stumps showed precise measurements when the fifth measurement
was analysed, Table 9. Using means of more than one measurement did not improve the precision; in some cases it made it worse.

Table 8. Interrater precision of stump volumes, ICC.

<table>
<thead>
<tr>
<th>Measure</th>
<th>ICC</th>
<th>95% CI Lower limit</th>
<th>95% CI Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure 5</td>
<td>0,76</td>
<td>0,47</td>
<td>0,94</td>
</tr>
<tr>
<td>Mean, measures 4 and 5</td>
<td>0,79</td>
<td>0,52</td>
<td>0,94</td>
</tr>
<tr>
<td>Mean, measures 3-5</td>
<td>0,79</td>
<td>0,52</td>
<td>0,94</td>
</tr>
<tr>
<td>Mean, measures 2-5</td>
<td>0,78</td>
<td>0,51</td>
<td>0,94</td>
</tr>
<tr>
<td>Mean, measures 1-5</td>
<td>0,78</td>
<td>0,51</td>
<td>0,94</td>
</tr>
</tbody>
</table>

ICCs less than 0,85 written in fat.

Table 9. Interrater precision of stump volumes, clinical criterion.

<table>
<thead>
<tr>
<th>Stump A</th>
<th>Stump B</th>
<th>Stump C</th>
<th>Stump D</th>
<th>Stump E</th>
<th>Stump F</th>
<th>Stump G</th>
<th>Stump H</th>
<th>Precise stumps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean volume, all measures</td>
<td>562,4</td>
<td>859,0</td>
<td>565,9</td>
<td>668,3</td>
<td>680,3</td>
<td>730,7</td>
<td>792,7</td>
<td>516,9</td>
</tr>
</tbody>
</table>

Measure 5

<table>
<thead>
<tr>
<th>Range (% of mean)</th>
<th>Precise measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>-12,4</td>
<td>-11,6</td>
</tr>
<tr>
<td>0/4</td>
<td>0/4</td>
</tr>
</tbody>
</table>

Mean, measures 4 and 5

<table>
<thead>
<tr>
<th>Range</th>
<th>Precise measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>-12,4</td>
<td>-10,9</td>
</tr>
<tr>
<td>11,2</td>
<td>15,5</td>
</tr>
<tr>
<td>0/4</td>
<td>1/4</td>
</tr>
</tbody>
</table>

Mean, measures 3-5

<table>
<thead>
<tr>
<th>Range</th>
<th>Precise measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>-12,6</td>
<td>-12,3</td>
</tr>
<tr>
<td>12,3</td>
<td>15,8</td>
</tr>
</tbody>
</table>

Mean, measures 2-5

<table>
<thead>
<tr>
<th>Range</th>
<th>Precise measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>-12,5</td>
<td>-14,3</td>
</tr>
<tr>
<td>11,2</td>
<td>16,1</td>
</tr>
</tbody>
</table>

Mean, measures 1-5

<table>
<thead>
<tr>
<th>Range</th>
<th>Precise measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>-11,0</td>
<td>-15,5</td>
</tr>
<tr>
<td>12,0</td>
<td>16,3</td>
</tr>
<tr>
<td>0/4</td>
<td>0/4</td>
</tr>
</tbody>
</table>

Errors larger than ±5% and imprecise measures written in fat.

Sources of error

The first factor, problems with axial measurements, was analysed by comparing the measured stump lengths. The lengths had a mean range of 0,7 cm within the same CPO, Table 10. The range between CPOs was wider, the mean range was 1,9 cm when analysing each CPO’s fifth measurement, and 3,0 cm when analysing all measurements, Table 11. The error limit was chosen to ±5% for the axial measurements, while this was the most conservative error limit calculated, Appendix III. 34,4% of all measurements were precise
when comparing different observers’ fifth measurement, Table 11. When comparing all measurements by the observers 38,1% of the measurements were precise enough.

### Table 10. Intrarater precision of stump lengths, expressed in ranges (cm).

<table>
<thead>
<tr>
<th>CPO 1</th>
<th>Stump A</th>
<th>Stump B</th>
<th>Stump C</th>
<th>Stump D</th>
<th>Stump E</th>
<th>Stump F</th>
<th>Stump G</th>
<th>Stump H</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,0</td>
<td>0,5</td>
<td>0,3</td>
<td>0,2</td>
<td>0,5</td>
<td>0,2</td>
<td>0,4</td>
<td>0,4</td>
<td>0,4</td>
<td></td>
</tr>
<tr>
<td>2,3</td>
<td>1,0</td>
<td>0,3</td>
<td>0,7</td>
<td>1,7</td>
<td>0,5</td>
<td>1,5</td>
<td>0,2</td>
<td>1,0</td>
<td></td>
</tr>
<tr>
<td>0,4</td>
<td>1,0</td>
<td>0,3</td>
<td>0,5</td>
<td>0,5</td>
<td>1,0</td>
<td>0,2</td>
<td>0,3</td>
<td>0,5</td>
<td></td>
</tr>
<tr>
<td>0,3</td>
<td>0,9</td>
<td>0,4</td>
<td>0,6</td>
<td>1,1</td>
<td>0,7</td>
<td>1,0</td>
<td>0,9</td>
<td>0,7</td>
<td></td>
</tr>
</tbody>
</table>

Mean all CPOs: 0,7

Ranges ≥1,0 cm written in fat.

### Table 11. Interrater precision of stump lengths.

<table>
<thead>
<tr>
<th>Mean lengths, all measures (cm)</th>
<th>Stump A</th>
<th>Stump B</th>
<th>Stump C</th>
<th>Stump D</th>
<th>Stump E</th>
<th>Stump F</th>
<th>Stump G</th>
<th>Stump H</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10,85</td>
<td>10,96</td>
<td>10,22</td>
<td>13,35</td>
<td>11,37</td>
<td>14,60</td>
<td>10,89</td>
<td>11,15</td>
<td>11,67</td>
</tr>
</tbody>
</table>

Measurement 5

| Precise measures (within ±5% of mean of all measures) | 2,0 | 3,3 | 1,0 | 0,7 | 2,7 | 1,3 | 2,8 | 1,3 | 1,9 |

All: All: 11/32 (34,4%)

All measurements

| Precise measures (within ±5% of mean of all measures) | 2,9 | 4,1 | 1,5 | 5,9 | 3,0 | 2,0 | 3,4 | 1,3 | 3,0 |

All: All: 61/160 (38,1%)

Ranges ≥1,0 cm written in fat.

The circumferential measurements taken at the constant mark had a mean range of 1,0 cm (3,6% of the mean circumference), Table 12. At proximal tibia and four cm from the distal end, the mean ranges were 1,2 cm (3,8%) and 2,4 cm (9,0%), respectively. The error limit for the circumferences was decided to ±2,5% while this was the most conservative limit calculated, Appendix III. This resulted in that 81,3% of the measurements were precise at the constant mark, Table 12. At proximal tibia and at four cm from distal end, the amount of precise measurements was 90,6% and 34,4%, respectively.

When circumferences taken at the same site, that is, at the constant mark, were compared, the mean range between observers was 1,0 cm and 81,3% of the measurements were precise enough, Table 12.
### Table 12. Interrater precision of circumferences at different sites, measurement 5.

<table>
<thead>
<tr>
<th>Stump</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stump</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circumference at constant mark seven cm distal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>26.5</td>
<td>35.1</td>
<td>28.8</td>
<td>26.0</td>
<td>29.1</td>
<td>27.4</td>
<td>33.1</td>
<td>25.8</td>
<td>28.9</td>
</tr>
<tr>
<td>Range</td>
<td>1,1</td>
<td>0,7</td>
<td>1,1</td>
<td>0,9</td>
<td>0,6</td>
<td>1,0</td>
<td>1,5</td>
<td>1,3</td>
<td>1,0</td>
</tr>
<tr>
<td>Range (% of mean)</td>
<td>4,1</td>
<td>2,0</td>
<td>3,8</td>
<td>3,5</td>
<td>2,1</td>
<td>3,6</td>
<td>4,5</td>
<td>5,0</td>
<td>3,6</td>
</tr>
<tr>
<td>Precise measures</td>
<td>4/4</td>
<td>4/4</td>
<td>3/4</td>
<td>4/4</td>
<td>4/4</td>
<td>0/4</td>
<td>3/4</td>
<td></td>
<td><strong>All: 26/32 (81,3%)</strong></td>
</tr>
<tr>
<td>(within ±2,5% of mean)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circumference at proximal tibia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>30.7</td>
<td>35.3</td>
<td>31.1</td>
<td>30.6</td>
<td>32.6</td>
<td>28.0</td>
<td>34.1</td>
<td>29.3</td>
<td>31.5</td>
</tr>
<tr>
<td>Range</td>
<td>2,3</td>
<td>1,0</td>
<td>0,4</td>
<td>1,2</td>
<td>1,7</td>
<td>0,8</td>
<td>1,5</td>
<td>0,6</td>
<td>1,2</td>
</tr>
<tr>
<td>Range (% of mean)</td>
<td>7,5</td>
<td>2,8</td>
<td>1,3</td>
<td>3,9</td>
<td>5,2</td>
<td>2,9</td>
<td>4,4</td>
<td>2,1</td>
<td>3,8</td>
</tr>
<tr>
<td>(within ±2,5% of mean)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circumference at four cm from distal end</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>25.1</td>
<td>32.4</td>
<td>26.2</td>
<td>21.2</td>
<td>27.1</td>
<td>24.2</td>
<td>31.2</td>
<td>23.4</td>
<td>26.3</td>
</tr>
<tr>
<td>Range</td>
<td>1,3</td>
<td>3,5</td>
<td>2,2</td>
<td>1,5</td>
<td>2,4</td>
<td>1,7</td>
<td>4,5</td>
<td>2,3</td>
<td>2,4</td>
</tr>
<tr>
<td>Range (% of mean)</td>
<td>5,2</td>
<td>10,8</td>
<td>8,4</td>
<td>7,1</td>
<td>8,9</td>
<td>7,0</td>
<td>14,4</td>
<td>9,9</td>
<td>9,0</td>
</tr>
<tr>
<td>Precise measures</td>
<td>3/4</td>
<td>1/4</td>
<td>1/4</td>
<td>3/4</td>
<td>2/4</td>
<td>1/4</td>
<td>0/4</td>
<td>0/4</td>
<td><strong>All: 11/32 (34,4%)</strong></td>
</tr>
<tr>
<td>(within ±2,5% of mean)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ranges ≥1,0 cm written in fat.

When using CAPOD the mean circumference taken at the base of the stump tip was below the acceptable limit (31,5 cm) for all models when taking the circumferences 5 cm from the tip end. When taking the circumferential measures 4 cm from the end, model 2 and 5 had mean circumferences below the acceptable limit (25,2 cm). The manual measurements showed that three models (1, 2, and 5) had too small mean circumferences 4 cm from the distal end.

As the height of the stump tip was decided to be 4 cm in the precision part of the study, the minimum circumference that could be allowed at the tip base was 25,2 cm. The mean of all 20 measurements on each stump was assumed to be the true value of each stump. Four of the eight stumps (A, D, F, and H) showed a mean circumference that was smaller than 25,2 cm. Table 12. When analysing the individual measurements two stumps (B and G) had all measurements larger than 25,2 cm.
DISCUSSION

Postoperative volume changes have shown to have an individual pattern for transtibial amputees (Golbranson et al. 1988). Therefore it is hard to give general recommendations about the time for fitting with definitive prosthesis. The advantage of following the volume changes of every patient is obvious, but the most methods used in scientific studies for volume determination are not suitable for clinical use. The method evaluated in the present study was simple and could potentially be used as a clinical routine.

Main results

The error limit used for the volumes was ±5% (5% is approximately one stocking). The evaluated method showed to be accurate in theory, when taking measures from CAPOD software, but not accurate on all stump models when measured manually. The intrarater precision was good when using the ICC but neither the intra- nor interrater precision was satisfying when using clinical criteria. The poor precision was caused by variation between estimations of stump tip heights and between circumferential measurements.

Accuracy

Theoretical accuracy of tip volume

The error of the tip volume was not constant for different models, varying between -1,41 and 5,75 cm$^3$. Neither was the error linear, varying between -0,90 and 5,64%. The errors could therefore not easily be corrected for. The formula approximated the stump tip as a segment of a sphere and this was obviously not accurate on all tip shapes. The error was as most 5,64% and the error should not have had big influence on the total stump volume.

Theoretical accuracy of stump volume

When assessing the accuracy for the whole models in the software, the error was less than ±5% for all models when taking the measurements five cm apart. The evaluated method overestimated the volume of all models, but the error was neither constant nor linear. The overestimation ranged from 4,14 to 25,72 cm$^3$ (0,60 to 4,38%) and could therefore not easily be corrected for. For five of the six models the overestimation of the volume was partly caused by an overestimation of the tip volume. This was probably not a big factor and was not true for model 4 where the tip volume was underestimated but the total volume was overestimated. The main reason for the overestimation of the volumes was rather that the circumferences of the stump models were not completely circular as assumed in the formula. This means that the volume between two circumferences on the stump actually could have been smaller than the calculated cut cone even though the circumferences were correctly estimated.

Practical accuracy of stump volume

The analysis of the practical accuracy showed that the systematic error could not easily be corrected for, while it was neither constant nor linear. When the measurements were taken
five cm apart the error varied between -24,06 and 127,55 cm$^3$ (1,84 and 12,50%). The variation of the errors was also considerable when the circumferences where taken four cm apart, the errors ranged from -28,09 to 121,54 cm$^3$ (-2,15 to 11,92%). For the two models measured with three cm between the circumferences the error also varied: 85,36 and 122,55 cm$^3$ (7,10 and 12,01%).

When measuring the models manually the errors were greater than in theory, the only cases when the errors were smaller for manual measurements were for model 1 (five and four cm between the measurements) and for model 2 (four cm between the measurements). These comparisons should be interpreted with caution while the scanned models were modified a little in the software and where therefore no longer identical to the models measured manually.

A limitation of the manual measurements was that the distance between the circumferences, $h$, was not at right angle to the circumferences while the stumps usually had a coned shape. This introduced a mathematical error in the manual measurements, which was not present in the measurements in the CAPOD software, which may have led to an overestimation of the volumes. Still, the volume was overestimated for only three of the six models both when the measurements were taken five and four cm apart. This does not completely support the theory.

There is a risk that several errors, for example the error of the tip volume, the lack of circularity of the circumferences and the mathematical error of $h$ previously discussed, influenced the results. In some cases the errors may have amplified each other and in other cases counteracted each other. The error of the tip volumes is hard to correct for. Although better approximations of the tip volume could be constructed mathematically, it is difficult to construct clinical criteria for how to choose between different approximation models of the tip. The lack of circularity of the circumferences cannot be improved without changing the shape of the stump with manual manipulations. While the stump is wished to keep in its natural shape during the measurements, manipulation of the shape is not a good alternative. When measuring recently amputated stumps, the oedema probably makes the circumferences of the stump more circular. The cone shape of the stumps introduced an error of the distance $h$. Estimating the angle of the cone shape of the stump and constructing a table for how $h$ should be modified in the calculations of stumps with different angels could potentially reduce this error. This would make the method more complicated and if the method can be made accurate and precise enough without this, it is preferable.

Surprisingly, there was not a constant improvement of the accuracy when the circumferences were measured closer to each other. For some models there was an improvement, but for others the accuracy got worse. The errors may have been caused by other factors than the distance between the circumferences and may therefore not have been improved when the circumferences where measured closer to each other. Such factors could be errors of the stump tip volume or lack of circularity of the circumferences. For four of the six models (66,7%) an acceptable accuracy could be obtained with five cm between the measurements. This is comparable to the results of Krouskop et al. (1979) where 11 of 16 transfemoral stump models (68,8%) had a volume within ±5% of a water immersion technique. In another study the method was evaluated on real transfemoral residual limbs (Krouskop et al. 1988). The method was compared to a water immersion technique and a contourgraph method. The volume estimations of the three methods were within 3-6% of each other for seven of the
eight stumps (87.5%). While the results were presented in diagrams only it was not possible to calculate the amount of measurements within the limits used in the present study.

For two models (model 4 and 5) neither five, four, nor three cm between the measurements resulted in an acceptable accuracy. Model 5, which had the largest error, had a very long and narrow shape. Model 4 had a curved shape, but on the other hand other models, which the method was accurate for, was curved too. While an acceptable accuracy could not be reached on all models in the present study the question can be raised if the method is appropriate for all kind of stump shapes. The models used in this study had a mature form and had not the typical swollen shape of recently amputated stumps. The question if the method is accurate on swollen stumps is still to be answered and should ideally be answered by taking measurements on recently amputated stumps, not on models.

**Precision**

**Intrarater precision**

The intrarater precision was very good for all CPOs when using the ICC (0.95–1.00) with the whole confidence interval above 0.85. When using clinical criteria the intrarater precision was acceptable for only one of the four CPOs.

**Interrater precision**

The ICC showed a much wider confidence interval for the interrater precision compared to the intrarater precision, overlapping the limit for an acceptable ICC, 0.85. This means that there is an uncertainty if the method is precise enough of not, when it comes to interrater precision. Using the average of more than one measurement improved the ICC, but only marginally. When using the clinical criteria the interrater precision was not satisfying, only three of the eight stumps (37.5%) fulfilled the criteria. Using the average of more than one measurement did not improve the results. Obviously there is a need to improve the interrater precision of the method.

In the study by Krouskop et al. (1979) the two persons were within ±5% of their mean volume for seven of the eight transfemoral stumps (87.5%). This is a better result than in the present study, but in Krouskop’s study transfemoral, not transtibial, stumps were evaluated and both persons performing the measurements used the same pen marks as sites for the measurements.

**Sources of error**

The measurements performed in the study were of the indirect type while the volumes of the stumps were calculated from measurements of circumferences and distances between the circumferences. The errors in the circumferences and distances will then propagate and influence the calculated volume. The two factors (distances and circumferences) were analysed separately in an attempt to find the sources of the errors. A big variation between assessed stump lengths was found in the analysis, and only 38.1% of the measurements were within the ±5% limit. This indicates that there were problems with the axial measurements. The difference in ranges and in amount of precise measurements (within the ±2.5% limit) was negligible between the circumferences at the constant mark (range 3.6% of mean circumference, 81.3% precise measurements) and at proximal tibia (range 3.8% of mean circumference, 90.6% precise measurements). The circumference at four cm from the distal
end, on the other hand, showed a much wider range (range 9.0%, 34.4% precise measurements). This indicates that the problems with axial measurements (assessed as stump length variations) were mainly due to problems with correctly assessing the height of the stump tips and not due to problems with palpating the proximal end of tibia. The precision of the method can probably be improved by using a longer rule with a setsquare end when assessing tip heights, which would decrease the subjective influence from the CPO. The weakness of this interpretation is that the circumference changes more rapidly when moving axially at the end of the stump compared to moving axially at the proximal stump. The greater variations between the circumferences at the tip may therefore not reflect greater variations between the axial measurements, but reflect a stump shape that makes the circumferential measurements vary more when the axial measurements vary.

When comparing circumferences taken by different CPOs at the same site the mean range between observers was 1.0 cm and 81.3% of the measurements were precise enough. This indicates that also the circumferential measurements were sources of errors, although this may not have been the main source of error. In some studies a spring-loaded handle has been attached to the measuring-tape to increase the precision of the circumferential measurements by keeping the tension of the tape constant in all measurements (Harrelson et al. 1998). This is unfortunately not possible when using a metal circumference rule, as used in the present study. Development of a flexible measuring-tape easy to disinfect would make it possible to use it on stumps with unhealed wounds. A spring-loaded handle could then be attached to increase the precision of the circumferential measurements.

The results showed that the approximation of the stump tip as a segment of a sphere was problematic. All models had too small circumferences at the tip base when taking the measurement 5 cm from the tip end in CAPOD. When taking the measure 4 cm from the end two models had too narrow tips. The problems persisted in the practical accuracy; three models did not have circumferences large enough. In the assessment of precision half of the stumps were too narrow distally, when analysing the mean values of the circumferences. Only stump B and G had all measurements above the limit, 25.2 cm. Both stump B and G had a rounded shape, Figure 9. The result questions if the method could be used on all types of stumps. The method seems to be more suitable for volume assessments of recently amputated stumps, which have a more rounded shape distally due to the postoperative oedema. If the method is to be used on narrow stumps another approximation has to be used for the volume of the stump tip. Taking the most distal circumference measure closer to the distal end would decrease the minimum acceptable circumference and perhaps make the method useful on stumps with a more narrow shape, but taking measures close to the tip is hard due to the curved shape of the tip.

The measurement errors were further analysed according to the error factors from the theory section, that is, errors of measurement are mainly caused by three factors: the observer, the instrument, and the study subject/object (Hulley and Cummings 1988).

The CPOs may have handled the measuring devices in different ways and had a different level of clinical experience. This can be improved by giving the CPOs instructions on how to measure and training in the measuring procedure. In the present study all CPOs were given written instructions and a form to fill in, Appendix I and II. A weakness was that the CPOs did not have any training in the procedure before the study. There was no attempt made to keep the CPOs blind to their own measurements, and they sometimes helped each other with
filling out the forms. There is a risk that the CPOs consciously or unconsciously may have modified the numbers they filled in their forms.

The measuring instruments may also have caused errors of measurement. Every wooden rule was measured with a calliper in an attempt to estimate the accuracy. No rule exceeded 20.03 cm, when measuring the 20 cm mark on the rule. Although the method for this estimation had great weaknesses the results indicate that the errors of the wooden rules were insignificant. The errors of the metal circumference rules were probably insignificant compared to other sources of error. The influence of temperature and moisture on the rule and circumference rule was not estimated but was probably also of negligible magnitude. A longer wooden rule had been a better choice while some stumps are longer than 20 cm. A longer rule would have made it possible to measure the tip heights without moving the rule from the first position along the tibial crest. A metal circumference rule was used throughout the study which is stiffer than the flexible measuring-tapes used by CPOs clinically, and therefore it does not follow the shape of the stump as good as a flexible measuring-tape. This may introduce errors in the circumferential measurements but the metal rule is easier to hold so the circumference is measured in one plane. Other advantages of the circumference rule are that it can be disinfected and is easier to use when measuring close to the stump tip.

The study subjects may also have caused variations in the estimated volumes. While the measuring session was time consuming there is a risk that the stump volumes may have fluctuated during the session. In the instructions the CPOs were requested to ask to patient to sit with extended knee and muscles relaxed during the measurements. Contraction of muscles has shown to increase the stump volume (Lilja et al. 1999) and a flexed knee may increase the circumference at the proximal stump.

**Clinical implications**

The method evaluated cannot be used in clinical practice because of its too weak precision. Further development of the method might solve this. The development work should focus on evaluation of the method’s accuracy on real stumps and increasing the precision, especially the intrarater precision. The precision can probably be improved by using a longer rule with a set square end. If this work is successful it would result in an inexpensive and simple method for volume estimations that could be used in clinical practice postoperatively as well as in scientific studies estimating volumes postoperatively or later in the rehabilitation.

Other methods could potentially also be used to measure the postoperative volume of amputees, but the method evaluated in this study has advantages over other methods for volume assessment. Both CAD/CAM systems and water immersion techniques are not possible to use when the patient is lying down, which is possible with the evaluated method. This is an advantage while many patients are in poor physical condition after the amputation. The CAD/CAM and water immersion apparatus may also be cumbersome to move and therefore demand that the patient is transported to the apparatus. The evaluated method does not require cumbersome equipment and the measurements can therefore be taken wherever the patient is situated. The method is also easy to learn, while the measurements taken are familiar to the clinically active CPO. A disadvantage of the method is that it is based on indirect measurements, where the stump volume has to be calculated from the measurements taken on the stump. Although direct measurements are very common at orthopaedic workshops, indirect measurements are never performed manually. Performing these calculations may seem complicated to some CPOs and some may therefore meet the method
with skepticism. On the other hand, both CAD/CAM system and water immersion technique also require training of the personal performing the measurements. In this study the volumes were calculated with a programmable calculator or a simple program written in Excel. The extra work was therefore neither complicated nor very time consuming.

In the clinical work other factors also influence the decision of when to fit the patient with prosthesis. Such factors may be the economy of the clinic, the time available for the local CPO, the physical and mental condition of the patient, etcetera.

The results of the study question the intrarater precision of circumferential measurements made on transtibial stumps, at least when the present circumference rule is used. The average range between measurements made by four different CPOs was 1.0 cm, although the circumferences were measured at the same site. These kinds of measurements are made very frequently at orthopedic workshops. If one wish to increase the intrarater precision a spring-loaded handle can be attached to the measuring-tape to standardise the measuring procedure, but this is only possible when using flexible measuring-tapes.

It is the author’s impression that there is a lack of awareness of errors of measurement at orthopaedic workshops. Although many CPOs are aware that there are variations between measurements when performed by different persons, the magnitude of the errors is not often known. There are many different types of measurements performed by CPOs and errors are bound to all these measurements. The amount of soft tissue on the stump, the shape of the stump, the temperature of the stump, etcetera, are relatively subjectively judged and may be influenced by gross errors, for example variations between different CPOs. Even measurements experienced more “objective”, as circumferences and lengths of residual limbs, are burdened with errors.

The problem is not solely how to investigate the problems of interest (for example errors of measurement), but how to spread the results to the CPOs at the clinics. Although errors of measurements taken with a measuring-tape have been investigated (Harrelson et al. 1998), the results are probably not well known in clinical practice. It is the author’s belief that a presentation at an annual conference is a good way to spread the results to the professionals in a country. An article published in a scientific paper does not reach all professionals, but still is a good way to communicate the results to an international audience.

**Methodological considerations and limitations of the study**

In the introduction several questions were asked to illustrate the meaning of external and internal accuracy, respectively. The questions shall now be repeated and answered to analyse the accuracy of the present study.

**External accuracy**

The external accuracy is concerned with the questions:

1. *Did the study plan answer the research question?* The aim of the study was to evaluate accuracy and precision of a specific method to assess stump volume. The accuracy was evaluated by theoretical comparisons inside CAPOD software and by measuring stump models. No real stumps were measured, which is a big weakness. Measuring rigid models is not the same as measuring soft residual limbs, and real stumps should
therefore preferable be used. This limits the possibility to generalize the results to real residual limbs. The study plan had potential to correctly assess the precision of the method. The precision was evaluated on real stumps, which gives a greater possibility to generalize the results outside the study. A weakness was that the accuracy and precision were evaluated separately, so the precision was unknown when assessing accuracy, and vice versa.

2. *Was what was stated as “truth” in the study plan possible to generalize to the universe outside the study?* The answer of this question is dependent on the answers of question 1, 3, 4, and 5. Except for what is discussed under these questions, the CPOs were not blinded to their own or to the other CPOs’ measurements. This is a factor that limits the ability to generalize the “truth” of the study.

3. *Did the intended sample reflect the accessible population the sample was drawn from?* No systematic selection of the subjects was used, and the author had not access to any information of the patients not participating. Therefore, it cannot be said for sure that the intended sample reflected the accessible population.

4. *Did the accessible population reflect the target population?* The accessible population was amputees in the Jönköping city area in Sweden. Sweden is a western country where the most amputations are performed on elderly people with vascular diseases (with or without diabetes). They cannot represent the world population of transtibial amputees that include many mine victims. The stump shape of a mine victim may be different from the stump shape after a planned amputation. Therefore the results cannot for certain be transferred to this kind of patients.

5. *Were the variables intended to measure good measures of the phenomena of interest?* No statistical tests were used for the comparisons inside the CAPOD software. Only one “measurement” was performed per volume. Additional measurements would simply have consisted of trying to measure the volume between the same coordinates on the stump as in previous measurements. The accuracy of the manual measurements was estimated with Student’s t-test, which is a widely accepted way to investigate if there is any difference between two populations. Two different criteria were used to estimate precision: an ICC of at least 0.85 and clinical criteria. Neither the ICC nor the clinical criteria gave an acceptable interrater precision. The intrarater precision, on the other hand, showed very good results when using the ICC, but only one CPO fulfilled the clinical criteria. This gives rise to the question which evaluation method is the most useful in this situation. The ICC can be used as an estimate of precision in almost any context but has the weakness that it does not show what the results mean in clinical terms. Therefore clinical criteria may be preferable to use when they are easy to identify.

**Internal accuracy**

The internal accuracy is concerned with the questions:

1. *Was the actual study performed in accordance with the study plan?* The accuracy part of the study was performed in close accordance to what was intended in the study plan. In the precision part of the study the ambition was to use five CPOs, but only four persons were used in reality. This did not influence the results much except for
that the confidence interval of the interrater precision, measured with the ICC, became wide (0.47-0.94). Another step away from the study plan was that one CPO occasionally used a second wooden rule hold at right angle with the first, as a help to estimate the tip heights. This is a source of error while it may have weakened the interrater precision and the intrarater precision for that CPO (CPO 3).

2. \textit{Did the study results support what was stated as “truth” in the study?} This is a question to the answered by the reader of this thesis. The author has described the present study in detail; how it was performed and its limitations. The results, their interpretations, and the sources of the errors have also been clearly described, but still the reader may find the interpretations made by the author disputable.

3. \textit{Did the study subjects reflect the intended sample?} The intention was to contact transtibial amputees in the Jönköping city area, which also was done. No systematic selection of the subject was performed, but this was never the intention. The study subjects therefore reflected the intended sample.

4. \textit{Did the measurements performed in the study measure the intended variables good enough, keeping errors of measurement small?} When assessing interrater precision, a randomised order is preferable to avoid variations due to learning effects when the person performing the measurements gets familiar with the method. A randomized order was introduced in this study but had to be left when the session was time consuming (about 2,5 hours) and some patients needed to leave earlier. This limits the internal accuracy of the study. If, for example, all CPOs measure the same stump as the first one and all measure another stump as the last one, the consistency between the measurements may be smaller for the last stump because of the increased experience of the method. The precision can also be worse for the last stump because the CPOs may be tired. With the long time from first to last measurement there is also a risk that the stumps may have varied in volume during the measuring session.

\textbf{Other methodological considerations}

The evaluated method approximated the stump volume as consisting of simple geometric shapes (cones and sphere segments). Neither the measurements nor the mathematics were then complicated. Still, the calculations were too time consuming to be performed manually and a programmable calculator or a computer was needed. Maybe more accurate models of the stump volume could be constructed using complicated mathematics (for example splines), without making the measuring procedure more complicated. This would result in a method that is both easy to use and gives accurate results, while the complicated part of the work is performed by a computer.

The results showed that it was preferable to take the most distal circumference at four cm compared to five cm from the end of the tip. Other distances could have been used, but taking the measurement too close to the stump end is complicated due to the shape of the stump. Another problem with measuring close to the tip is that although the error in the estimation of the tip may decrease, the error above the measurement may increase because the whole tip shape is not included in the approximation of the tip.

This study has only treated fitting with definitive prosthesis as a problem of stump volume. In reality, socket fit does not solely depend on socket volume but also on socket shape. If the
shrinkage postoperatively is equal around the whole stump, some stockings can compensate for the shrinkage. On the other hand, if the shrinkage is very local, fitting problems may arise without big volume changes. The shape changes of the stump are thus a factor of interest to investigate. The disadvantage with shape is that it is much harder to quantify than volume. Volume is not a perfect measure of socket fit, the question is if volume is a measure good enough of socket fit postoperatively.

Future studies
Further development and evaluation of the method is needed. The author’s suggestion is that a study is constructed where a number of CPOs measure a number of transtibial stumps with postoperative oedema. The stumps should also be measured with a reference method, for example water immersion technique. In this way accuracy and precision could be estimated in the same time, which is an advantage. A simpler approach was used in this study, with the disadvantage that the interrater precision was not known when the author measured the models and the true stump volumes were not known when assessing precision. The order of the measurements should be randomised and the CPOs blinded to each other’s measurements. The measurement sites should not be marked direct on the stump but on a removable tape so every CPO has to identify the sites for the measurements. Accuracy and intra- and interrater precision should be evaluated with clinical criteria and systematic errors corrected for if possible. The measuring equipment should be a rule, about 30 cm, and an instrument to measure circumferences with, possible to disinfect. A metal circumference rule could be used, but a flexible measuring-tape is preferable while a spring-loaded handle then can be attached to increases the precision of the circumferential measurements.
CONCLUSIONS

The aim of the study was to evaluate accuracy and intra- and interrater precision of a method to calculate stump volume from circumferential measurements. Error limits were a volume error of ±5% (5% is one stocking) and an ICC of at least 0.85. It was found that the method, in theory, is accurate enough. In reality the method was not accurate on all stump types, but this part was done on stump models and evaluation of accuracy on real stumps is needed. The approximation of the stump tip was not valid for narrow stumps.

Although the intrarater precision was satisfying when using the ICC, both the intra- and interrater precision was unsatisfying when using clinical criteria. Therefore, further development of the method is needed before it can be recommended for clinical use. The weak precision was caused by variation between estimated tip heights and between circumferential measurements. Using a longer rule with a setsquare end for measuring the tip heights of the stumps is recommended for improvement of the precision. If a flexible measuring-tape possible to disinfect is developed, a spring-loaded handle could be used to increase the precision of the circumferential measurements.
ACKNOWLEDGEMENTS

The author would like to attribute thanks to Magnus Lilja, CPO, PhD, Anton Johannesson, CPO, Kenth Lund, CPO, CPOs at Team Ortopedteknik in Jönköping (Karin, Paul, Martin, and Anna), and all patients participating in the study.
REFERENCES


Appendix I, Instructions for the measurements

The patient sits with his knee in the most extended position possible. Ask the patient to (except for the extension) relax the stump.

Begin with putting a long adhesive tape on the tibial crest. Then place the rule so it begins by the proximal, anterior edge of tibia. Let the rule lay in the imagined line from mid-patella to the tibial end.

Put a mark on the tape 4 cm from the distal end of the stump. Put another mark by the proximal end of tibia. Then go distal and put a mark every five cm. When less than five cm is left to the first mark you put, record the distance left to the mark and fill in the last but one box on the form.

Then record the circumferences by the marks and fill in the form from the bottom. Struck any remaining empty boxes and circles.

**Finally, take a circumferential measure by the mark.**

Between every measurement the tape you put on is removed (not the little piece of tape). Ask the patient to let the stump down between the measurements.

The measurement is repeated five times on every patient.

All measures are rounded off to the closest millimetre.

Notice that the circumference rule is not to be tightened but just to lie against the skin.

**Randomised order of measurements:**

CPO 1: stump H, F, G, C, D, B, A, E  
CPO 2: stump B, D, G, E, H, A, F, C  
CPO 3: stump B, A, D, E, C, H, G, F  
CPO 4: stump A, H, F, B, G, C, E, D  
CPO 5: stump D, F, G, B, C, E, H, A
Appendix II, Measuring form

Patient:______________
Date:_______________

4

Patient:______________
Date:_______________

4

Patient:______________
Date:_______________

4

Patient:______________
Date:_______________

4
Appendix III, Calculations

Error limits for cut cone

The volume of a cut cone is calculated as:

\[ V = h \cdot \frac{C_j^2 + C_k^2 + C_jC_k}{12\pi} \]  
[10]

The three partial derivatives are taken from equation [10]:

\[ \frac{\partial V}{\partial h} = \frac{C_j^2 + C_k^2 + C_jC_k}{12\pi} \]  
[11]

\[ \frac{\partial V}{\partial C_j} = \frac{h(2C_j + C_k)}{12\pi} \]  
[12]

\[ \frac{\partial V}{\partial C_k} = \frac{h(C_j + 2C_k)}{12\pi} \]  
[13]

The propagation of error can then be approximated with the formula:

\[ \Delta V = \frac{\partial V}{\partial h} \Delta h + \frac{\partial V}{\partial C_j} \Delta C_j + \frac{\partial V}{\partial C_k} \Delta C_k \]

A. Assume that \( C_j \) and \( C_k \) are without error, that is, \( \Delta C_j = \Delta C_k = 0 \). The influence of the error \( \Delta h \) can then be calculated as:

\[ \Delta V = \frac{\partial V}{\partial h} \Delta h \]

where the partial derivative is substituted by the expression from equation [11].

\[ \Delta V = \frac{C_j^2 + C_k^2 + C_jC_k}{12\pi} \Delta h \]

The expression is divided with \( V \) to find the relative error of \( V \), \( \varepsilon_v \):

\[ \varepsilon_v = \frac{\Delta V}{V} = \frac{C_j^2 + C_k^2 + C_jC_k}{12\pi \cdot V} \Delta h \]
\[ \Delta h \text{ is replaced with the expression:} \]
\[ \varepsilon_h = \frac{\Delta h}{h} \]

\[ \Delta h = \varepsilon_h \cdot h \]

The relative error of the volume is then found with the expression:
\[ \varepsilon_v = \frac{(C_j^3 + C_k^3 + C_j C_k)\varepsilon_h \cdot h}{12\pi V} \]

V is substituted by the expression in equation [10]:
\[ \varepsilon_v = \frac{(C_j^3 + C_k^3 + C_j C_k)\varepsilon_h \cdot h}{12\pi \left( h \cdot \frac{C_j^3 + C_k^3 + C_j C_k}{12\pi} \right)} \]

\[ \varepsilon_v = \varepsilon_h \]

Thus, a relative error of h will result in a relative error of the volume of the same magnitude. When the relative error of the volume is allowed to be ±5\%, that is, \( \varepsilon_v = \pm 0.05 \), the relative error of h is allowed to take a maximum value of ±5\%.

B. Assume that the circumferences are the sources of error. Assume further that h is without error and the relative errors of \( C_j \) and \( C_k \) are equal, that is:

\[ \Delta h = 0 \]
\[ \varepsilon_j = \varepsilon_k = \varepsilon \]

\[ \frac{\Delta C_j}{C_j} = \frac{\Delta C_k}{C_k} = \varepsilon \]

The propagation of the error is then calculated with the formula:
\[ \Delta V = \frac{\partial V}{\partial C_j} \Delta C_j + \frac{\partial V}{\partial C_k} \Delta C_k \]

which is divided by V to find the relative error of the volume:
\[ \varepsilon = \frac{\Delta V}{V} = \frac{\frac{\partial V}{\partial C_j} \Delta C_j + \frac{\partial V}{\partial C_k} \Delta C_k}{V} \]

The partial derivatives are replaced by the expressions in equation [12] and [13].

\[ \varepsilon = \frac{\frac{h(2C_j + C_i)}{12\pi} \Delta C_j + \frac{h(C_j + 2C_i)}{12\pi} \Delta C_k}{V} \]

\(\Delta C_j\) and \(\Delta C_k\) are substituted by

\[ \Delta C_j = C_j \cdot \varepsilon \]

\[ \Delta C_k = C_k \cdot \varepsilon \]

\[ \varepsilon = \frac{\frac{h(2C_j + C_i)}{12\pi} C_j \cdot \varepsilon + \frac{h(C_j + 2C_i)}{12\pi} C_i \cdot \varepsilon}{V} \]

\[ \varepsilon = \frac{2h \cdot \varepsilon \cdot (C_j^2 + C_k^2 + C_i C_j)}{12\pi V} \]

\(V\) is substituted by the expression from equation [10]:

\[ \varepsilon = \frac{2h \cdot \varepsilon \cdot (C_j^2 + C_k^2 + C_i C_j)}{12\pi \cdot h \cdot (C_j^2 + C_k^2 + C_i C_j)} \]

\[ \varepsilon = 2\varepsilon \]

\[ \varepsilon = \frac{\varepsilon}{2} \]

Thus, the relative error of the circumferences is half of the relative error of the volume. If the limit for the relative error of the volume is ±5%, the relative error of the circumferences cannot be allowed to exceed 2.5%, if the error of \(h\) is zero and the relative errors of the circumferences \(C_j\) and \(C_k\) are equal:

\[ \varepsilon = \frac{0.05}{2} = 0.025 = 2.5\% \]
Error limits for sphere segment

The volume of a sphere segment is calculated as:

\[ V = \frac{\pi t^3}{6} + \frac{C^2 t}{8\pi} = \frac{8\pi^2 t^3 + 6C^2 t}{48\pi} \]  \[ \text{[14]} \]

The partial derivatives of equation [14] are:

\[ \frac{\partial V}{\partial t} = \frac{\pi t^2}{2} + \frac{C^2}{8\pi} = \frac{4\pi^2 t^2 + C^2}{8\pi} \]  \[ \text{[15]} \]

\[ \frac{\partial V}{\partial C} = \frac{C^* t}{4\pi} \]  \[ \text{[16]} \]

The propagation of error is approximated with the formula:

\[ \Delta V = \frac{\partial V}{\partial t} * \Delta t + \frac{\partial V}{\partial C} * \Delta C \]  \[ \text{[17]} \]

A. Assume that C is without error, that is \( \Delta C = 0 \). Equation [17] then reduces to:

\[ \Delta V = \frac{\partial V}{\partial t} * \Delta t \]

which is divided by V to find the relative error of V, \( \varepsilon_v \):

\[ \varepsilon_v = \frac{\Delta V}{V} = \frac{\frac{\partial V}{\partial t} * \Delta t}{V} \]

\( \Delta t \) is replaced by:

\[ \varepsilon_t = \frac{\Delta t}{t} \]

\( \Delta t = \varepsilon_t * t \)

\[ \varepsilon_v = \frac{\frac{\partial V}{\partial t} * \varepsilon_t * t}{V} \]
The partial derivative is substituted by the expression in equation [15]:

\[ \varepsilon_t = \frac{4\pi^2 t^2 + C^2}{8\pi} \cdot \varepsilon_t \cdot \frac{t}{V} \]

V is substituted by the expression in [14]:

\[ \varepsilon_t = \frac{4\pi^2 t^2 + C^2}{8\pi} \cdot \varepsilon_t \cdot \frac{48\pi}{8\pi^2 t^2 + 6C^2 t} \]

The relative error of t can then be found with the equation:

\[ \varepsilon_t = \frac{\varepsilon_t (3C^2 + 4\pi^2 t^2)}{C^2 + 4\pi^2 t^2} \]

Assume that \( t = 4 \) and \( \varepsilon_v = \pm 0.05 \), as used in the precision evaluation of the method. All circumferences measured by the CPOs were in the interval 20-35 cm. \( \varepsilon_t \) will then be:

- \( C = 20 \rightarrow \varepsilon_t = \pm 8.88\% \)
- \( C = 25 \rightarrow \varepsilon_t = \pm 9.97\% \)
- \( C = 30 \rightarrow \varepsilon_t = \pm 10.88\% \)
- \( C = 35 \rightarrow \varepsilon_t = \pm 11.60\% \)

All above calculated error limits for \( t \) are larger than the previously calculated error limit for \( h \), \( \pm 5\% \). \( \pm 5\% \) is therefore chosen as the error limit for the axial measurements, that is, for \( h \) and \( t \).

B. Assume that \( t \) is without error, that is \( \Delta t = 0 \). Equation [17] then reduces to:

\[ \Delta V = \frac{\partial V}{\partial C} \cdot \Delta C \]

The expression is divided with \( V \) to obtain the relative error of the volume, \( \varepsilon_v \):

\[ \varepsilon_v = \frac{\Delta V}{V} = \frac{\partial V}{\partial C} \cdot \frac{\Delta C}{V} \]

The partial derivative is replaced by the expression in equation [16]:

\[ \varepsilon_v = \frac{C^* t * \Delta C}{4\pi V} \]
\[ \Delta C \text{ is replaced by:} \]
\[ \varepsilon_c = \frac{\Delta C}{C} \]
\[ \Delta C = \varepsilon_c \cdot C \]

and \( V \) is replaced by the expression in equation [14]. The relative error of \( V \) is then:

\[ \varepsilon_v = \frac{C^2 \cdot t \cdot \varepsilon_c \cdot C \cdot 48\pi}{4\pi \cdot 8\pi^2 t^3 + 6C^2 t} \]

\[ \varepsilon_v = \frac{C^2 \cdot t \cdot \varepsilon_c \cdot 24}{4\pi^2 t^3 + 3C^2 t} \]

\[ \varepsilon_v = \frac{6C^2 \varepsilon_c}{4\pi^2 t^2 + 3C^2} \]

The relative error of the circumference can then be found with the equation:

\[ \varepsilon_c = \frac{\varepsilon_v (4\pi^2 t^2 + 3C^2)}{6C^2} \]

Assume that \( t = 4 \) and \( \varepsilon_v = \pm 0.05 \). \( \varepsilon_c \) will then be (for the same circumferences as previously):

\[ C = 20 \rightarrow \varepsilon_c = \pm 3.82\% \]
\[ C = 25 \rightarrow \varepsilon_c = \pm 3.34\% \]
\[ C = 30 \rightarrow \varepsilon_c = \pm 3.08\% \]
\[ C = 35 \rightarrow \varepsilon_c = \pm 2.93\% \]

All error limits above are larger than the previously calculated error limit for circumferential measurements, \( \pm 2.5\% \). \( \pm 2.5\% \) is therefore chosen as the error limit for the circumferential measurements.