

RELIABILITY AND LIMITS OF AGREEMENT OF CIRCUMFERENTIAL, WATER DISPLACEMENT, AND OPTOELECTRONIC VOLUMETRY IN THE MEASUREMENT OF UPPER LIMB LYMPHEDEMA

T. Deltombe, J. Jamart, S. Recloux, C. Legrand, N. Vandebroek, S. Theys, P. Hanson

Departments of Physical Medicine and Rehabilitation (TD,SR,CL,NV,ST,PH) and Biostatistics and Medical Documentation (JJ), Université Catholique de Louvain, Cliniques Universitaires de Mont-Godinne, Yvoir, Belgium

ABSTRACT

We conducted a reliability comparison study to determine the intrarater and interrater reliability and the limits of agreement of the volume estimated by circumferential measurements using the frustum sign method and the disk model method, by water displacement volumetry, and by infrared optoelectronic volumetry in the assessment of upper limb lymphedema. Thirty women with lymphedema following axillary lymph node dissection surgery for breast cancer surgery were enrolled. In each patient, the volumes of the upper limbs were estimated by three physical therapists using circumference measurements, water displacement and optoelectronic volumetry. One of the physical therapists performed each measure twice. Intraclass correlation coefficients (ICCs), relative differences, and limits of agreement were determined. Intrarater and interrater reliability ICCs ranged from 0.94 to 1. Intrarater relative differences were 1.9% for the disk model method, 3.2% for the frustum sign model method, 2.9% for water displacement volumetry, and 1.5% for optoelectronic volumetry. Intrarater reliability was always better than interrater, except for the optoelectronic method. Intrarater and interrater limits of agreement were calculated for each technique. The disk model method

and optoelectronic volumetry had better reliability than the frustum sign method and water displacement volumetry, which is usually considered to be the gold standard. In terms of low-cost, simplicity, and reliability, we recommend the disk model method as the method of choice in clinical practice. Since intrarater reliability was always better than interrater reliability (except for optoelectronic volumetry), patients should therefore, ideally, always be evaluated by the same therapist. Additionally, the limits of agreement must be taken into account when determining the response of a patient to treatment.

Keywords: Lymphedema, reliability, circumferential measurements, water displacement, volumetry

The reported incidence of lymphedema after breast cancer surgery ranges from 6% to 30% (1). The extent of axillary operation and exposure to axillary radiotherapy are independent risk factors for the development of lymphedema (2). Lymphedema can cause heaviness, discomfort, impaired function, subcutaneous tissue infection, and an unsatisfactory appearance. Lymphedema treatment programs provided by physical therapists are multimodal, including a combination of manual lymphatic drainage, pneumatic compression, bandaging, sleeve wearing, and

education (3,4). In our practice, we need to use rapid, simple, non-invasive and reliable measurement tools for determination of the volume of upper limb lymphedema (4,5). Currently, water displacement volumetry is considered to be the gold standard (6,7). Other commonly used techniques include estimation of the volume from measurements of the limb circumference with tape, and infrared optoelectronic volumetry (8). Skin tonometry (9), bioelectrical impedance (10), and computed tomography (11) have also been proposed. There is no report in the literature comparing the reliability of the four most frequently used methods for the assessment of upper limb volume. Specific objectives of the study were: (1) to assess the intrarater and interrater reliability of the volume estimated from circumferential measurements, water displacement volumetry, and optoelectronic volumetry; (2) to compare the intrarater and interrater reliability of these different methods; (3) and to determine the limits of agreement of each method.

METHODS

Subjects

Thirty women with chronic arm lymphedema secondary to unilateral breast cancer treatment volunteered for the study. The diagnostic criterion for lymphedema was a clinically evident involvement of the arm. All subjects were included from a rehabilitation program provided in our department including manual lymphatic drainage, pneumatic compression, and bandaging. Written informed consent was obtained from all subjects and the study was approved by the Ethics Committee of our institution. Fifteen subjects had been treated for breast cancer on the right side and 15 on the left side. All subjects had undergone axillary lymph node dissection, 27 had undergone total mastectomy, and 3 partial mastectomy. Eight subjects had undergone chemotherapy and all but one had previously had radiation therapy

to the breast and axillary nodes. Exclusion factors were a continuing neoplastic process, bilateral surgery, recent lymphangitis, and brachial plexitis. Age ranged from 46 to 79 years (63 ± 9 years, mean \pm SD), their body height ranged from 152 to 176 cm (162 ± 6 cm) and their body weight ranged from 48 to 103 Kg (74 ± 12 Kg). The median delay between surgery and appearance of lymphedema was 28 months (interquartile range: 12;64). The median delay between the appearance of lymphedema and the study was 126 months. The precipitating factor was unknown in 20 cases (67%), unusual effort in 4 cases (13%), both direct trauma and reconstructive surgery in 2 cases, and lymphangitis and orthopedic device in one case.

Measurement Procedures

All subjects were assessed on both arms (lymphedema side and contralateral side) at the beginning of the treatment. The procedure used was similar to our clinical practice. The evaluation of the arms included measurement of their circumference with a tape, and volume determination by water displacement volumetry and by infrared optoelectronic volumetry. Three physical therapists performed all measurements in a random order one after another on each patient. There was no additional wait time between the different examinations so volume should not have changed between successive measurements. The first therapist (PT1) performed each measure twice while the second (PT2) and third (PT3) performed each measure once in order to limit the duration of the experiment (the complete procedure took approximately 1 hour). PT1 was an undergraduate physical therapist student while PT2 and PT3 had experience in lymphedema measurement of 8 and 15 years, respectively. Before starting the study, the three physical therapists practiced the techniques together but they worked independently during the study. An independent examiner collected the data for statistical analysis.



Fig. 1. Circumferential measurement of the upper limb.

Circumferential measurements

For circumferential measurements, subjects were placed in a sitting position with forearms pronated (*Fig. 1*). A plastic tape measure with an accuracy of 1 mm was placed around the arm, always in direct contact with the skin but without excessive pressure. Measurements of arm circumference were made with the tape at the level of the metacarpal shaft for the hand, just below the ulnar styloid process for the wrist, and every 5 cm along the arm 20 cm distal and 20 cm proximal to the lateral epicondyle. The circumference was, therefore, measured in 12 different places. The marks were completely erased between each measurement performed by a different therapist.

The volume of the arm was calculated from the circumference measurements by two methods. The simplest method is the frustum sign model method (also called single truncated cone method), which uses physics to determine the volume of a truncated cone. The application of this formula involves measurement of only the lower (c) and the upper (C) circumferences of the arm. The circumferences chosen were located at the wrist for the lower circumference and 20 cm proximal to the lateral epicondyle for the

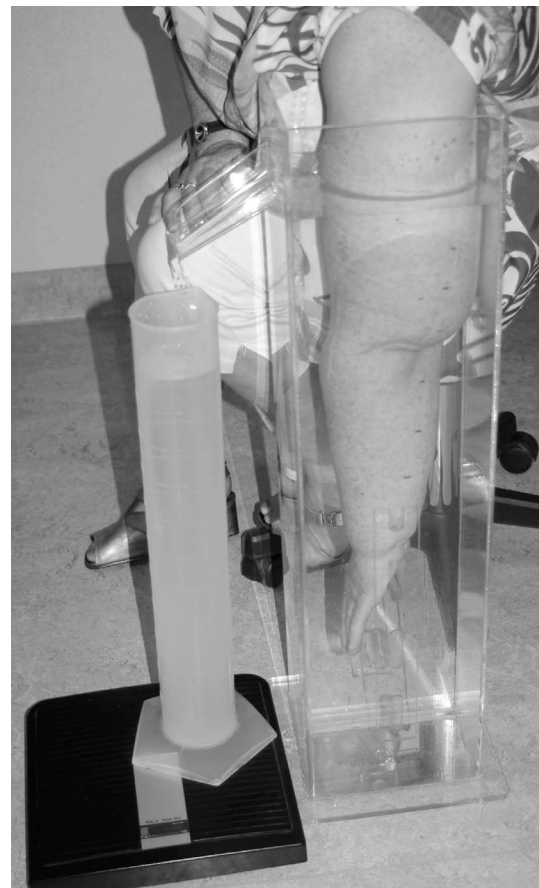


Fig. 2. Water displacement volumetry of the upper limb.

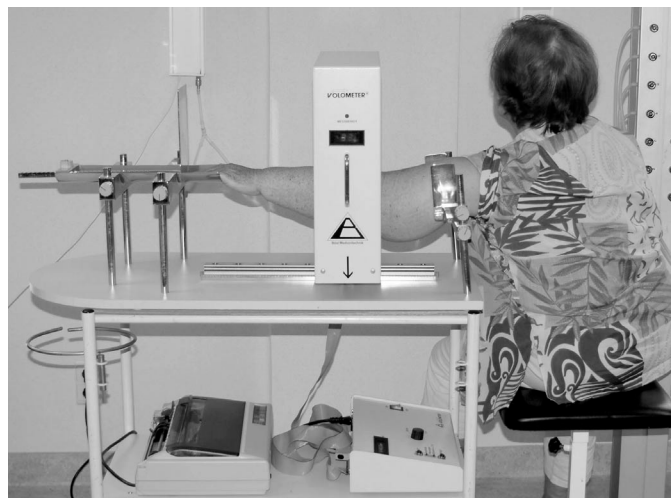


Fig. 3. Infrared optoelectronic measurement of the upper limb.

upper circumference. The formula is $V = h(C^2 + Cc + c^2)/12\pi$, where h is the distance between the upper and lower circumference (15). The second method is the disk model method (also called summed truncated cone method), which divides the arm into 10 disks with a height of 5 cm. The formula considers the base of the first disk as the circumference at the hand and subsequent disks at 5 cm intervals until 20 cm above the epicondyle. The volume is then calculated as the sum of the individual disk volumes using: $V = \sum (C^2 h/4\pi)$, where C is the circumference of the disk and h is the height of the disk (7).

Water displacement volumetry

Water displacement volumetry was performed with a volumetric device (Arm Volumeter®, Volumeters Unlimited, Phoenix, AZ 85032) filled with water at 20°C (Fig. 2). Subjects were seated and instructed to put their arm vertically with the elbow straight in the pronated position into the cylinder, up to a dermatographic skin mark located 15 cm below the acromion. The overflowing water was weighed with an electronic balance (Philips HR 2385/A Austria) which had an accuracy of 1 gr. We chose to weigh the water instead of measuring the volume of over-

flowing water because volume determination with graduated cylinders had lower accuracy (10 ml for a 250 ml graduated cylinder) than the electronic balance (accuracy of 1 gr). The kilogram weight was converted to milliliters using the standard conversion of 1kg = 1000 mL.

Infrared optoelectronic volumetry

The infrared optoelectronic volumetric measurement was performed with a commercially available device (Volometer®, Bösl Medizintechnik, Aachen, Germany). Subjects were seated and instructed to put their arm horizontally into the Volometer with the elbow straight in the pronated position with the tip of the middle finger placed down the device (Fig. 3). The Volometer consists of a horizontally movable frame equipped with infrared light emitters and receptors. The infra-red light beams are interrupted by the introduction of the arm into the frame. The shadow is captured on the X and Y-axis. By moving the frame along the long axis of the arm a measure is automatically performed every cm for a distance of 40 cm. The arm volume is then calculated from these measures.

Data Analysis

The reliability of measurements was studied using three methods. We computed the intraclass correlation coefficient (ICC) with 95% confidence interval in a two-way random model (subject and measurement). This parameter expresses the part of the measurement variability that is due to the subjects with respect to the total measurement variability due to the subject and the observer. We also computed the difference between two measurements, divided by their mean and multiplied by 100 to obtain what is called the relative difference. The intrarater relative difference is the difference between the two measurements performed by physical therapist PT1. The interrater relative difference is the difference between the measurements performed by physical therapist PT2 and PT3. These pairs of measurements were compared using the Wilcoxon signed rank test, which was also used to compare the intrarater and interrater differences of the various methods. Lastly, we determined the limits of agreement according to Bland and Altman (12). These values are estimated by the mean difference (bias) \pm two standard deviations of this difference. They represent the limits of reproducibility from the observed experiments which are to be considered as acceptable or not from a clinical viewpoint. Statistical analysis was performed with SPSS statistical software (SPSS Inc., Chicago, IL).

RESULTS

Table 1 lists the intrarater and interrater intra-class correlation coefficients (ICC), means and standard deviations of relative differences, and limits of agreement of the volumes estimated from circumferential measurements by the frustum sign and the disk model methods, from water displacement volumetry, and from optoelectronic volumetry. All techniques had satisfactory ICC, ranging from 0.937 to 0.997. The ICCs for intrarater and interrater reliability for the frustum sign

TABLE 1
Relative Difference, Intraclass Correlation Coefficient (ICC) and Limits of Agreement of Intrarater and Interrater Reliability of the Volume Estimated from the Circumferential Measurement by the Frustum Sign and the Disk Model Method, by Water Displacement Volumetry and by Infrared Optoelectronic Volumetry

	FRUSTUM SIGN		DISC MODEL		WATER		OPTO-ELECTRONIC	
	Intra	Inter	Intra	Inter	Intra	Inter	Intra	Inter
ICC	0.958	0.937	0.989	0.990	0.991	0.987	0.997	0.997
95% confidence interval	0.940-0.970	0.911-0.956	0.984-0.992	0.985-0.993	0.987-0.994	0.981-0.991	0.995-0.998	0.995-0.998
Relative difference (%)	3.2 \pm 4.6	4.8 \pm 3.2	1.9 \pm 2.9	3.1 \pm 2.2	2.9 \pm 2.9	4.5 \pm 3.8	1.5 \pm 1.4	1.7 \pm 1.6
Limits of agreements (ml)	\pm 348	\pm 470	\pm 286	\pm 332	\pm 252	\pm 327	\pm 114	\pm 112

TABLE 2
Limits of Agreement of the Intrarater and Interrater Circumferential Measurements (cm) at Each Level

	Intrarater	Interrater
Finger	± 0.917	± 1.336
Hand	± 0.747	± 1.308
Wrist	± 1.311	± 2.548
- 20 cm	± 1.300	± 1.72
- 15 cm	± 2.291	± 2.072
- 10 cm	± 1.920	± 1.864
- 5 cm	± 1.730	± 1.528
Elbow	± 1.539	± 1.585
+ 5 cm	± 1.697	± 1.885
+ 10 cm	± 1.785	± 1.931
+ 15 cm	± 1.877	± 2.241
+ 20 cm	± 1.997	± 2.05

method were 0.96 and 0.94, respectively. The ICCs for the intrarater and interrater reliability for the disc model method, water displacement volumetry, and optoelectronic volumetry were 0.99. The intrarater relative differences for the frustum sign method ($3.2 \pm 4.6\%$) and water displacement volumetry ($2.9 \pm 2.9\%$) were greater than for the disk model method ($1.9 \pm 2.9\%$) and optoelectronic volumetry ($1.5 \pm 1.4\%$) ($p < 0.001$). The interrater relative differences for the frustum sign method ($4.8 \pm 3.2\%$) and water displacement volumetry ($4.5 \pm 3.8\%$) were greater than for the disk model method ($3.1 \pm 2.2\%$) ($p < 0.001$ and $p = 0.003$, respectively). The optoelectronic volumetry had the lowest interrater relative differences ($1.7 \pm 1.6\%$) ($p < 0.001$). The intrarater relative differences were lower than interrater relative differences for the volume estimated from circumferential measurements ($p < 0.001$) and for water displacement volumetry ($p < 0.001$) while there was no significant difference for optoelectronic volumetry (NS). The intrarater limits of

agreement of each technique were smaller than the interrater limits of agreement for all techniques except optoelectronic volumetry. Table 2 lists the limits of agreement of the intrarater and interrater circumferential measurements at each level.

DISCUSSION

Precise determination of the volume of upper limb lymphedema has three important goals. The first is the early diagnosis of lymphedema enabling appropriate treatment before clinically significant lymphedema develops. All women who receive axillary lymph node dissection in the management of breast cancer are at increased risk of developing lymphedema and should be referred to a physical therapist experienced in lymphedema measurement in order to provide proper assessment and adequate information about lymphedema. This is of importance as we know that delayed intervention in reducing lymphedema may result in poor functional outcomes. The second goal is to follow the volumetric evolution of the lymphedema and monitor the effects of treatment. Many therapeutic modalities such as manual lymphatic drainage, pneumatic compression or bandaging are effective in reducing the volume of upper limb lymphedema but individual responses to these different treatments are often variable and largely unpredictable. Precise knowledge of the volume decrease subsequent to a given treatment (lymphatic drainage or compression for instance) may help to determine the most effective treatment for each patient in order to propose individual programs based on the objective results obtained by previous treatments. The last goal is to facilitate comparison of different treatments and methods applied to a large population of patients in order to objectively compare them.

The studies which have dealt with the reliability of the different methods for measurement of upper limb lymphedema after breast cancer surgery were performed

to evaluate whether water displacement volumetry, considered by some authors to be the “gold standard,” could be replaced by volume estimated from circumference measurements (6,13-15). The authors reported good correlation coefficients varying from 0.93 to 0.99 for comparisons of water displacement volumetry and the frustum sign and disk model method (6,13,14). A high correlation coefficient between two methods indicates that the volumes determined by these methods will change in the same direction with both methods. However, even if the correlation coefficients are high, there can be considerable discrepancy between methods (7). By determining the limits of agreement between two methods, all authors concluded that calculated and volumetric measurements are both reliable and closely related but do not agree with each other, and thus should not be used interchangeably (6,13,14,16). Sander et al even concluded that such an analysis cannot indicate which method is preferable (13). The term “gold standard” is, moreover, largely inappropriate as in practice the water displacement volumetry does not give the exact volume of the limb (although the best method for “true” volume). The validity of the different methods for the measurement of upper limb lymphedema has to be estimated by comparing each of these methods with the exact volume, not with one of the methods whatever its supposed performance. Thus, we have restricted our study to the assessment of the reliability of the different methods and not to the determination of the validity, which is necessarily biased by confounding a supposed gold standard with the reality. The aim of our study was, thus, not to compare the volume determined by different methods nor to determine whether one method can be replaced by another for volume determination, but to compare the reliability (by means of the relative differences) and to determine the limits of agreement of the different methods available for the measurement of upper limb lymphedema after breast cancer

treatment. We found good intrarater and interrater reliability for the different methods with ICCs ranging from 0.99 to 1. Only the frustum sign method had lower reliability with ICCs ranging from 0.94 to 0.96. Such results are consistent with previous studies, which report ICCs ranging from 0.96 to 0.99 (6,13-17). Based on the ICC results, Sander et al support the premise that different therapists can reliably obtain the measurements if they train together with an established protocol. However, even if ICCs are high, intrarater and interrater reliability expressed by means of the relative differences can reveal significant variation ranging from $1.5 \pm 1.4\%$ to $4.8 \pm 3.2\%$. By comparing the reliability of the different methods, we found that the intrarater reliability was always better than the interrater reliability except for optoelectronic volumetry, which had similar intrarater and interrater reliability. The better intrarater than interrater reliability is reinforced by the fact that the intrarater reliability was calculated from the measurements performed by an undergraduate physical therapist student with only limited experience in the methods while the interrater reliability was calculated from the measurements of the two very experienced physical therapists who had worked for 8 and 15 years, respectively, in our lymphedema unit. Therefore, except for optoelectronic volumetry, we argue that ideally a patient should always be evaluated by the same therapist even if he/she has limited experience. The intrarater reliability was significantly better with optoelectronic volumetry ($1.5 \pm 1.4\%$) and the disk model method ($1.9 \pm 2.9\%$) than with water displacement volumetry ($2.9 \pm 2.9\%$) and the frustum sign method ($3.2 \pm 4.6\%$).

Thus, we propose the disk model method as the best method to assess upper limb lymphedema in routine office practice especially when performed by the same therapist. Girth measurement is a simple and low cost technique that can be used irrespective of skin condition and requires minimal technology. Only a tape and a personal computer

are needed for volume determination from circumferential measurements (and the calculations can even be done by hand if necessary). It took approximately 10 minutes to perform bilateral evaluation of the upper limbs. Furthermore, circumferential measurements provide interesting segmental information on the location and extent of lymphedema. Indeed, volume determination, whatever the method used, gives a global assessment of the upper arm. Circumferential measurements, without volume calculation, are often the most concrete way of identifying the location of a lymphedema in the upper limb and of assessing the effects of treatment on each part of it. Technical limitations, which may induce variability in the measures, are differences in traction applied on the tape, in positioning of the tape, and in landmark location. The infrared optoelectronic volumetry has the best reliability and is the only method that provided no difference between intrarater and interrater reliability suggesting that different therapists can use it on the same patient. It is also the quickest method, taking only a few seconds, provides segmental information (every centimeter), and can be used when skin lesions are present. The main limitation of the device is that the measurement, performed only to a height of 40 cm, provides no information about lymphedema located above the elbow. The cost of the device limits its use to specialized centers and clinical trials.

Water displacement volumetry is considered by many authors as the best method for volume measurement of the lower and upper limbs (6,7). All the studies undertaken to compare the volume determined by different techniques have used water volumetry as reference. Recently, some authors have raised doubts regarding the reliability of water displacement volumetry. Sander et al reported that other methods than water displacement volumetry had smaller measurement errors for both the arm and the hand (13). We found that water displacement volumetry provided lower intrarater and

interrater reliability than the disk model method and optoelectronic volumetry and that interrater reliability was lower than intrarater reliability. Such poor reliability can be explained by the fact that our study was conducted in a clinical situation where time is an issue, that positioning the upper limb strictly vertically in the device can be difficult, and that water may not be correctly collected and weighed. In addition, water displacement volumetry cannot be used where there are skin lesions and is not suitable in patients in the immediate postoperative period. There is also a risk of cross infection if the system is not sterilized between patients at risk. Changing or refilling the water and allowing time for it to equilibrate also makes the technique time consuming.

In contrast with previous studies, we did not use the limits of agreements to determine the level of agreement between two measurements methods but to determine the clinical significance of variation of a given method as initially developed by Bland and Altman (12). Indeed, determining the limits of agreement of a measurement method is the only way to answer this critical question: "How can we be sure that changes in measure observed on a given patient are due to the treatment and not to an error in the measure caused by an inappropriate method or usage?" The intrarater and interrater limits of agreement of each technique (*Table 2*) must be taken into account before determining the response of a patient to treatment. The limits of agreement of the intrarater and interrater circumferential measurements at each level were approximately ± 1 cm for the hand and ± 2 cm for the forearm and arm. So we can be certain that modifications of the circumference are due to the treatment only if they are greater than 1 cm for the hand and 2 cm for the forearm and arm. In conclusion, we have demonstrated that intrarater reliability is usually better than interrater reliability for the different methods used in volume determination of the upper limb lymphedema. Therefore, ideally, a patient should always be

evaluated by the same therapist. By comparing the reliability of the different methods, we have shown that the disk model method and optoelectronic volumetry are the most reliable methods. In view of its low cost and simplicity, we recommend the disk model method should be used in clinical practice.

CLINICAL MESSAGES

- 1) Compared to water displacement volumetry, the circumferential measurement has better reliability. This simple and low cost technique should be considered as the method for lymphedema measurement in clinical practice.
- 2) Ideally, a patient should always be evaluated by the same therapist even if the latter has limited experience.

ACKNOWLEDGMENT

The authors thank Mr. Bob Brew for his help. No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated. The Volumeter was purchased at regular price.

REFERENCES

1. Petrek, JA, MC Heelan: Incidence of breast carcinoma-related lymphedema. *Cancer* 83 (1998), 2776-2781.
2. Kissin, MW: Risk of lymphoedema following the treatment of breast cancer. *Br. J. Surg.* 73 (1986), 580-584.
3. Bunce, IH, BR Mirolo, JM Hennessy, et al: Post-mastectomy lymphoedema treatment and measurement. *Med. J. Aust.* 161(1994), 125-128.
4. Megens, A, SR Harris: Physical therapist management of lymphedema following treatment for breast cancer: A critical review of its effectiveness. *Phys. Ther.* 78 (1998), 1302-1311.
5. Gerber, LH: A review of measures of lymphedema. *Cancer* 83 (1998), 2803-2804.
6. Megens, AM, SR Harris, C Kim-Sing, et al: Measurement of upper extremity volume in women after axillary dissection for breast cancer. *Arch. Phys. Med. Rehabil.* 82 (2001), 1639-1644.
7. Kaulesar Sukul, D, P den Hoed, E Johannes, et al: Direct and indirect methods for the quantification of leg volume: Comparison between water displacement volumetry, the disk model method and the frustum sign model method, using the correlation coefficient and the limits of agreements. *J. Biomed. Eng.* 15 (1993), 477-480.
8. Tierney, S, M Aslam, K Rennie, et al: Infrared optoelectronic volumetry, the ideal way to measure limb volume. *Eur. J. Vasc. Endovasc. Surg.* 12 (1996), 412-417.
9. Chen, HC, BM O'Brien, JJ Pribaz, et al: The use of tonometry in the assessment of upper extremity lymphoedema. *Br. J. Plast. Surg.* 41 (1988), 399-402.
10. Cornish, BH, IH Bunce, LC Ward, et al: Bioelectrical impedance for monitoring the efficacy of lymphoedema treatment programmes. *Breast Cancer Res. Treat.* 38 (1996), 169-176.
11. Collins, CD, PS Mortimer, H D'Ettore, et al: Computed tomography in the assessment of response to limb compression in unilateral lymphoedema. *Clin. Radiol.* 50 (1995), 541-544.
12. Bland, JM, DG Altman: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* (1986), 307-310.
13. Sander, AP, NM Hajer, K Hemenway, et al: Upper extremity volume measurements in women with lymphedema: A comparison of measurements obtained via water displacement with geometrically determined volume. *Phys. Ther.* 82 (2002), 1201-1212.
14. Karges, JR, BE Mark, SJ Stikeleather, et al: Concurrent validity of upper-extremity volume estimates: Comparison of calculated volume derived from girth measurements and water displacement volume. *Phys. Ther.* 83 (2003), 134-145.
15. Sagen, A, R Karesen, MA Risberg: The reliability of a simplified water displacement instrument: A method for measuring arm volume. *Arch. Phys. Med. Rehab.* 86 (2005), 86- 83.
16. Taylor, R, UW Jayasinghe, L Koelmeyer, et al: Reliability and validity of arm volume measurements for assessment of lymphedema. *Phys. Ther.* 86 (2006), 205-214.
17. Meijer, RS, JS Rietman, JHB Geertzen, et al: Validity and intra- and interobserver reliability of an indirect volume measurements in patients with upper extremity lymphedema. *Lymphology* 37 (2004), 127-133.

Th. Deltombe, MD
Department of Physical Medicine
and Rehabilitation
Cliniques Universitaires de Mont-Godinne
UCL (Université Catholique de Louvain)
B-5530 Yvoir, Belgium
Tel: 32 81 42 21 11 Fax : 32 81 42 37 79
E-Mail : Thierry.Deltombe@uclouvain.be