National Defence / Défense nationale

Canada

[redacted]

Defence and Civil INSTITUTE OF ENVIRONMENTAL MEDICINE

INSTITUT DE MEDECINE ENVIRONNEMENTALE pour la défense

1133 Sheppard Avenue West, PO Box 2000, North York, Ontario, Canada M3M 3B9
Tel. (416) 635-2000 Fax. (416) 635-2104

January 1999 DCIEM CR1999-041

ANTHROPOMETRIC SURVEY OF THE LAND FORCES: MEASUREMENT ERROR ANALYSIS

by

Francis Forest*
André Chamberland*
Jean-Michel Billette*
Pierre Meunier+

(DCIEM +)

Les Consultants Génicom Inc. *
3400 Boul. de Maisonneuve Ouest, Montréal, Québec, Canada, H3Z 3B8

Project Managers:
Robert Carrier and André Chamberland
(514) 931-3000

PWGSC Contract No. W7711-6-7350

on behalf of
DEPARTMENT OF NATIONAL DEFENCE
as represented by
Defence and Civil Institute of Environmental Medicine+
1133 Sheppard Avenue West
North York, Ontario, Canada
M3M 3B9

DCIEM Scientific Authority:
Pierre Meunier
(416) 635-2093

TABLE OF CONTENTS

EXECUTIVE SUMMARY I

INTRODUCTION 1

METHODS 2

RESULTS 5

CONCLUSIONS 17

REFERENCES 18

Executive Summary

The Canadian Forces carried out an anthropometric survey for which the results were published in 1998. In order to evaluate the precision and accuracy of the measurements performed in the survey, an error analysis study was performed in parallel with the main survey. Precision, which is the difference between repeated measurements during the survey, was first analysed. . For this portion of the analysis, each of the last seven participants of the day were assigned to a particular station in order to be reassessed for a given set of variables after the whole session at the end of each day. Each subject had the same landmarks.

Accuracy, which is the difference between the measurements during the survey and the true value, was also monitored throughout the survey to avoid the emergence of systematic bias. To this end, 4 individuals from the survey's team of observers (the landmarkers) were selected. The "true value" was obtained through averaging the values obtained by a group of trained measurers for each of the anthropometric variables. Each of the 4 individuals was re-landmarked and re-measured at regular intervals throughout the survey, i.e. over a period of 2 months. This enabled the survey team to control the quality of the measurements by addressing landmarking and measurement biases.

The comparison of these two data sets with the US Army's MADs (Mean Absolute Difference between two measurements) and AOEs (Allowable Observer Error) was performed for the 140 survey measurements. The results show the measurement errors of the Land Forces survey to be comparable in magnitude to those obtained in the 1988 US Army survey.

This study highlights the occurrence of errors in measurements examined twice by the same examiner, errors resulting of the landmarking as well as variations during the time frame (up to two months) for the same measurement of the same subject.

Introduction

The Canadian Forces recently published the results of an anthropometric survey of the Land Forces (Chamberland et al, 1998). A total of 140 direct measurements and 60 indirect measurements were taken on 465 men and 243 women selected by means of random sampling. Summary statistics were produced for these 200 variables : means, standard deviations, standard errors, coefficients of relative variation, lowest and highest values, kurtosis, skewness, percentiles and frequency tables. These statistics were calculated separately for men and women.

Data concerning the size and shape of the user population are essential for the design of clothing and personal equipment as well as for workspace design and evaluation. The primary purpose of the survey was to satisfy to the particular needs of the Canadian Land Forces by providing data as accurately and precisely as possible.

Given that no anthropometric survey is devoid of error of measurement, the problem that is faced is not the elimination of error but its control. And this control can be achieved by giving the observers adequate training and by using a proper error accounting method for each direct measurement taken. By doing so, the data user knows exactly what degree of confidence he or she can attribute to every measurement(s) needed.

The question of anthropometric error has already been the focus of several studies in the past (see Dahlberg, 1940 for example). Many authors have used different ways of quantifying error margins, each method having its pros and cons. Russel et al (1994) provides a good overview of the many indexes that have been used in the past, with a critique of each of them.

When two or more successive measurements of the same variable are taken on the same subject, one can use, to compute reliability estimates, the Student's paired t test (PTT), the mean absolute difference (MAD), the intraclass correlation coefficient (ICC) or a multivariate analysis of variance (MANOVA) that allows the analyst to discriminate among the many sources of error.

In our case, we used the method developed by Gordon et al (1989), who performed the same kind of survey for the US Army in 1988. We also adopted the same approach with regard to the choice of the variables, the anthropometric techniques specific to each single measurement, as well as the way of evaluating our training's quality from the first workshops to the fieldwork's reliability tests.

Following Gordon and Bradtmiller's (1992) procedures, we decided to organize the observers in 7 different teams composed of a male and a female observer. Each team was responsible for about 20 measurements (one seventh of the 140 measurements). Survey participants were measured by an observer of the same sex. We also decided, in accordance with these authors, to meet or exceed the limits of the acceptable measurement error that were established for each of the survey variables and to maintain and control the quality of the measurements in the field.

Methods

It is generally accepted that there is no way to fully eliminate error of measurement in anthropometry, however scholars usually agree on the means of keeping it within certain acceptable limits. While preparing their survey, Gordon et al (1989) carried out a study whose goal was to estimate reliability limits one should not exceed between two values of the same measure. The allowable observer error limits were established using information available in the literature, data obtained from other surveys, and through ad-hoc tests using 4 expert observers and ten subjects (Gordon et al 1989). Intensive sessions were carried out where two pairs of two participants were measured in the morning and another two pairs in the afternoon, thus allowing 4 comparisons of a same subject examined twice in a short time span. It is noteworthy to mention that the participant's landmarks were the same for all measurement sessions.

The absolute difference was calculated for each of those 4 comparisons and the highest value concerning every measurement was fixed as the allowed observer error (AOE). The choice of this criterion is justified by the argument that the observers were long-trained specialists. This so-called allowed error is nevertheless acceptable in so far as it reflects the technical limitations one has to face while taking anthropometric measurements. This does not infer that, below this threshold, the ergonomic use of these data will be devoid of major repercussions. Furthermore, the allowed error does not take into account the possible differences that may result from different landmarkings.

Those allowed errors are the ones we dealt with to control the measurement precision from the first moments of the training period until the end of our data gathering on the field.

While training, this control allowed us to divide up our staff by attributing to every observer the particular set of measurements where he or she obtained the highest score. For this purpose, a part of the training period was spent assessing each observer's score for 10 subjects measured 4 times/day during two weeks. These scores allowed us to estimate the average values of the 140 measurements of every subject. We then computed, for each observer and for each measurement, the average deviation as well as the occurrence of sessions where this deviation fell within the range of allowed errors established by Gordon et al (1989). These results allowed us to compose two possible teams for each station. The first one was made up of the two observers (one of each sex) who had the lowest mean of mean deviations (3.89) and the second one was composed of the two observers who had a maximum observations within the range of the allowed error (251). Finally, the selected combination was the one with the best overall score, given that the mean deviation outcome was of greater value than the occurrence of acceptable measures (Combination 1). Table 1 shows the optimal combination of male and female teams determined on the basis of mean deviations and number of measurements within allowable limits.

Table 1: Team performance summaries

Combinations	Stations	Teams	Mean deviations (mm)	Number of measures within allowable error
1	A	M2/F4	3.44 / 3.21	250 / 400
	B	M1/F2	5.13 / 5.08	240 / 400
	C	M5/F7	3.60 / 3.34	250 / 400
	D	M4/F9	3.52 / 3.17	230 / 400
	E	M3/F1	4.19 / 4.13	220 / 400
	F	M8/F6	2.37 / 1.93	218 / 400
	G	M6/F3	4.97 / 6.55	270 / 400
		Mean	3.89 / 3.91	240 / 400
2	A	M6/F6	3.72 / 3.21	240 / 400
	B	M8/F3	5.65 / 5.08	280 / 400
	C	M5/F7	3.60 / 3.34	250 / 400
	D	M4/F1	3.76 / 3.17	230 / 400
	E	M1/F4	5.20 / 4.13	230 / 400
	F	M2/F2	2.46 / 1.93	250 / 400
	G	M3/F9	5.59 / 6.55	280 / 400
		Mean	4.29 / 3.92	251 / 400

Once the assignment was done, each of the combination 1 teams was trained to measure the variables it would be responsible for in the survey. The purpose was to reach the optimal quality standards by keeping the measured values as close as possible to the already computed mean values. Each observer's performance was reviewed daily in order to bring the appropriate corrections if necessary.

While gathering the data on the field, Gordon et al (1989) re-examined a certain number of participants for a particular series of measurements in order to compute the mean absolute difference between the two observers composing a team. These values were then compared to the allowed error for every single measure in order to detect possible bias that could have surfaced during the fieldwork.

This option was unfortunately unavailable to us due to the composition of our specialized teams (one male and one female); men were measured by men and women by women. Hence, we had to resort to two different solutions in order to maintain the quality of the observations.

The first solution involved re-measuring participants while keeping the same observer. Twice or thrice a week, each of the last seven participants of the day were assigned to a particular station in order to be reassessed for a given set of variables. That kind of test was repeated ten times during the survey, thus providing the team supervisor with a means of detecting possible biases in method, accuracy and precision.

This test, however, was not informative with respect to the emergence of a systematic bias over time. To prevent that possibility from occurring, we used a second solution which consisted of measuring once a week the same group of 4 subjects (our markers) about whom we had previous standard anthropometric information.

A total of 17 full sessions were performed on those four markers (2 of each sex) over the few weeks of fieldwork. In every case, the absolute difference between the observed measure and the average of the previously computed results for each variable was calculated, thus providing an estimation of an observer's deviation from the norm. This score, however, differs significantly from the MAD and the AOE in that it represents the difference between a measurement and a mean rather than the difference between two measurements. These results should be multiplied by two to be directly comparable to the former. Thus, this difference between the results of this test (T-2) shown in table 3 trough 11, should be taken into account.

The results of the survey will be discussed in the next section. The AOE and the MAD from the U.S. Army survey will be compared with those obtained in both our tests (test #1: ten participants measured twice; test #2 : 17 comparisons with the standard values for 4 subjects).

The results of these three different tests will be interpreted with respect to their specific characteristics, as shown in table 2. One can see that the computed MAD for the U.S. Army reflect an interobserver difference while the landmarks and the subjects are the same. In our first test, however, the MADs involve intraobserver variation. In the case of our second test, the standard value for each subject was fixed at the very beginning of the project. We can then figure out that the two major sources of discrepancy from the norm are: 1-variation in landmarking and 2-the intravariation of biological attributes over time (2months). Weight, for example, greatly fluctuated during this period.

Table 2 : Comparison of key elements of three observer error studies.

Study	**Landmarks**	**Observer**	**Subjects**	**Time difference**
USA	same	different	same	same day
Test 1	same	same	same	same day
Test 2	Different	same	norms	2 months

Results

All of our measurements were classified according to the specific type (heights, breadths, circumferences, etc...) they belong to in order to detect the homogeneity in terms of measurement error. The MAD is calculated by using the following formula:

$$MAD = \sum_{i=1}^{n} \left|(x_{i1} - x_{i2})\right| / n$$

where x_{i1} is the value of one measurement taken on subject *i* by the first observer and x_{i2} the value corresponding to the same measure taken on subject *i* by the second observer. All those differences are expressed in absolute values and computed before being divided by the total number of subjects in order to obtain an estimate of their mean.

The 140 direct measurements have been divided up between 9 types. Each table mentions the variable names as well as the number to which it corresponds in our final report (Chamberland et al, 1998). The first column shows the AOE and MAD from the American survey (Gordon, 1989). In the US Army, the MADs were listed separately for men and women, however those reported in our tables consist of the combined (male and female) scores because our own samples (n=10 and n=17) were not large enough to allow us to separate by gender. Our own MAD follow as well as the percent scores (MAD*100/ AOE) for each test (U.S.%, T1% and T2%). Finally for each specific metric, means were computed.

Thus in table 3, one can see that the average AOE for the 20 heights is 6.7 mm whereas the mean MAD (observer error) in the U.S. Army survey was 2.5 mm for the same measures. In our study, the intraobserver average MAD was 2.1 mm while the 17-comparisons-with-the-norms mean MAD was 2.7mm. These three results, when expressed in percent scores with regard to AOE, are respectively 39%, 34% and 47%. It is then logical to infer that the MADs obtained for the subjects' heights are below their respective AOE, especially in the cases of the U.S. Army and of our first test. The lower value obtained for the latter can be attributed to the fact that it represents intraobserver error while U.S. Army data reflect interobserver differences.

Table 3 : MAD for Heights

Variable	AOE (mm)	USA (mm)	T-1 (mm)	T-2 (mm)	US%	T1%	T2%
002-Stature	10	2.8	2.4	2	28	24	20
011-Cervical Height	7	2.3	1.7	2.2	33	24	31
012-Neck Height Lateral	7	2.4	1.8	2.1	35	26	30
013-Acromial Height	7	3.5	2.2	3.7	50	31	52
014-Suprasternal Height	5	2.8	1.5	2.4	55	30	48
015-Axilla Height	10	3.2	2.3	3.6	32	23	36
016-Chest Height	11	2.2	2.3	2.7	20	21	24
017-Tenth Rib Height	5	2.2	2.6	2.1	43	52	42
018-Waist Height, n.i.	4	2.2	1.0	3.2	55	25	80
019-Iliocristale Height	5	2.5	2.0	4.4	50	40	88
020-Waist Height Omphalion	7	2.9	1.8	1.7	41	26	24
006-Trochanteric Height	7	1.7	2.5	26	25	36	37
021-Buttock Height	7	1.6	2.6	24	23	37	34
022-Crotch Height	10	4.8	3.7	37	48	37	37
082-Gluteal Furrow Height	6	2.2	4.1	2.0	36	68	33
102-Wrist Height	11	5.6	2.6	2.5	51	24	23
083-Knee Height Midpatella	6	2.7	2.8	3.1	45	47	52
084-Lat.Fem.Epic.Height	3	1.4	2.1	2.2	46	70	73
085-Calf Height	3	1.1	0.9	4.5	38	30	150
086-Lateral Malleolus Height	3	0.7	0.1	1.0	22	3	33
Mean	**6.7**	**2.5**	**2.1**	**2.7**	**39**	**34**	**47**

AOE = Allowed Observer Error; USA = United States Army; US% = USA / AOE
T-1 = Test 1 (n=10) T1% = T-1 / AOE; T-2 = Test 2 (n=17); T2% = T-2 / AOE

Our test #2, on the other hand, shows MADs that are generally larger. Hence, one must keep in mind that these results involve differences in landmarking that are not present in both other tests. Let's give as examples the high T2% score concerning the waist height (natural indentation; 79%), the iliocristale height (87%) and the calf height (149%). In this last case, the marker must, before any measurement is taken, identify the point where the calf circumference reaches its highest value. Identifying twice, visually, the same point involves more variability in landmarking than the estimation of a relatively fixed (same landmarking in the USA and T-1 tests) point located 340 mm above the ground level. Estimating the waist height as well as the iliocristale height can also be difficult if the subject tend towards endomorphic profile, thus causing variability in landmarking.

The results for breadths are quite similar (table 4). The U.S. Army MAD represent on average 47% of the AOE while in our first test, average MAD represents 31% of the AOE. On the other hand, this score is 52% for T2%.

Table 4 : MAD for Breadths

Variable	AOE (mm)	USA (mm)	T-1 (mm)	T-2 (mm)	US%	T1%	T2%
047-Biacromial Breadth	8	3.6	2.1	2.9	45	27	37
049-Bideltoid Breadth	8	3.5	2.3	4.3	43	28	53
050-Chest Breadth	8	3.4	1.9	3.1	42	24	39
051-Thelion-Thelion Bread	10	3.7	1.8	2.1	37	18	21
052-Waist Breadth	6	2.6	1.9	4.6	43	32	77
053-Bispinous Breadth	3	2.0	1.5	4.2	68	50	139
054-Hip Breadth	7	2.5	1.1	2.5	35	16	36
055-Chest Depth	4	3.0	1.6	2.3	74	40	59
056-Waist Depth	8	2.7	2.8	2.6	34	35	32
057-Buttock Depth	8	3.7	3.2	3	47	40	37
Mean	**7.0**	**3.1**	**2.0**	**3.2**	**47**	**31**	**52**

AOE = Allowed Observer Error; USA = United States Army; US% = USA / AOE
T-1 = Test 1 (n=10) T1% = T-1 / AOE; T-2 = Test 2 (n=17); T2% = T-2 / AOE

While waist breadth errors may originate from landmarking as well as from instrument manipulation, those relative to the bispinous breadth (139%) are essentially landmarking dependent, since the observer does not have to apply pressure on the calliper to obtain the measure.

The 27 circumferences listed in table 5 yielded good results, the means being respectively 3.7, 1.6 and 2.8 mm for each study. The highest percent scores were found in our test #2 for the thigh circumference (61%), the knee circumference (97%) and the ankle circumference (69%). Differences in landmarking continued to contribute, to a large extent, to the overall error.

Table 5: MAD for Circumferences

Variable	AOE (mm)	USA (mm)	T-1 (mm)	T-2 (mm)	US%	T1%	T2%
145-Head Circumference	5	1.0	1.4	1.7	20	29	34
146-Neck Circumference	6	2.9	1.6	1.6	49	27	27
147-Neck Circumference Base	13	3.4	1.0	2.9	26	8	22
060-Shoulder Circumference	22	5.5	1.9	2.6	25	9	12
061-Chest Circumference Scye	15	6.6	1.7	3.1	44	11	21
062-Chest Circumference	15	6.6	1.7	2.1	44	11	14
063-Chest Circ. Below Breast	16	6.8	1.8	5.4	42	11	34
064-Waist Circumference n.i.	11	4.7	2.5	4.4	43	23	40
065-Waist Cricumference, omphalion	12	5.3	2.2	2.0	45	18	17
066-Buttock Circumference	12	4.3	2.2	5.6	36	18	47
089-Thigh Circumference	6	3.1	2.5	3.7	52	42	61
090-Lower Thigh Circmference	4	2.3	1.7	2.4	56	43	59
091-Knee Circumference	4	2.7	2.1	3.9	67	52	97
092-Calf Circumference	5	1.5	1.2	1.9	29	24	38
093-Ankle Circumference	4	1.5	1.3	3.0	37	32	69
099-Ball of Foot Circumference	4	2.0	1.0	1.6	49	25	38
100-Heel Ankle Circmference	6	4.5	0.6	2.3	76	10	38
113-Scye Circumference	13	5.6	1.4	4.7	43	11	36
114-Axillary Arm Circumference	8	3.2	1.2	2.1	40	15	26
115-Biceps Circumference, Relaxed	6	-----	2.3	1.3	-----	38	21
116-Biceps Circumference, Contracted	6	2.7	2.1	1.7	45	35	28
117-Elbow Circumference	4	1.5	0.9	1.6	37	22	40
118-Forearm Circumference, Relaxed	5	-----	1.5	1.7	-----	30	33
119-Forearm Circumference, Contracted	5	2.7	2.3	2.4	54	46	47
120-Wrist Circumference	5	1.3	0.1	0.7	25	2	14
121-Hand Circumference	4	0.8	0.5	0.7	19	13	16
069-VerticalTrunk Circumference	22	9.6	1.5	10.1	43	7	45
Mean	**8.8**	**3.7**	**1.6**	**2.8**	**42**	**23**	**36**

AOE = Allowed Observer Error; USA = United States Army; US% = USA / AOE
T-1 = Test 1 (n=10) T1% = T-1 / AOE; T-2 = Test 2 (n=17); T2% = T-2 / AOE

Table 6 shows results obtained for the 25 contours measured during the survey. The relative MADs are similar to those of the previous tables (51%, 35% and 53%). A great number of contours exceeded 75% in the case of T2%: scye depth, waist-back length (omphalion), waist front length (omphalion), shoulder length, interscye I and radiale-stylion length.

Table 6 : MAD for Contours

Variable	AOE (mm)	USA (mm)	T-1 (mm)	T-2 (mm)	US%	T1%	T2%
027-Scye Depth	4	2.6	1.3	4.3	66	32	108
028-Waist Back Length, n.i	5	3.0	3.0	3.4	61	60	73
029-Waist Back Length, o.	5	3.2	3.0	5.1	64	60	102
030-Waist Front Length, n.i.	5	2.8	1.8	2.3	57	36	55
031-Waist Front Length, o.	5	3.5	2.2	4.5	70	44	90
068-Strap Length	12	5.9	2.8	2.8	49	23	23
032-Neck-Thelion Length	8	3.1	1.1	2.7	39	14	34
046-Shoulder Length	3	2.2	1.2	2.7	73	40	90
122-Sleeve Outseam	6	3.4	0.8	0.6	57	13	10
058-Interscye I	10	6.3	4.6	9.8	63	46	98
059-Interscye II	13	5.7	4.0	6.6	44	31	50
125-Sleeve Spine-Scye	11	5.3	2.0	2.2	49	18	20
126-Sleeve Spine-Elbow	10	4.3	2.9	4.7	43	29	46
127-Sleeve Spine-Wrist	9	5.1	2.8	2.5	57	31	27
044-Waist n.i. Waist o. Length	3	1.2	2.7	1.7	39	90	57
045-Waist-Hip Length	6	1.8	1.9	3.7	29	32	61
072-Crotch Length n.i.	16	8.0	1.8	5.4	50	11	33
073-Cortch Length Posterior, n.i.	11	6.2	3.5	7.6	57	32	69
075-Crotch Length o.	18	7.4	12.4	4.0	41	69	22
076-Crotch Length Posterior o.	11	5.8	2.7	3.8	52	25	35
104-Acromion-Radiale Length	4	1.6	1.1	3.2	40	28	80
108-Radiale-Stylion Length	6	2.6	1.2	2.1	43	20	35
109-Shoulder-Elbow Length	6	2.1	1.8	2.0	34	30	33
112-Forearm-Hand Length	4	1.9	1.5	1.3	49	38	31
123-Sleeve Inseam Length	6	-----	1.3	2.8	-----	22	47
Mean	**7.9**	**4.0**	**2.6**	**3.7**	**51**	**35**	**53**

AOE = Allowed Observer Error; USA = United States Army; US% = USA / AOE
T-1 = Test 1 (n=10) T1% = T-1 / AOE; T-2 = Test 2 (n=17); T2% = T-2 / AOE

Let's consider the case of the waist back length as an example. The allowable error, according to Gordon et al (1989) is 5mm. The mean error in the U.S. Army survey was 3.2 mm with 2 different observers and 3.0 mm in our test #1 with a single observer. In our comparison with the norms, the mean distance between one particular value and the standard mean was 5.1 mm. This measurement was about 460 mm on average, for both genders in the Canadian Army. It is therefore clear that, within this range of 5 mm, instrument manipulation as well as the subject's exact position (the degree of back erection) can play a major role. Once again, differences in landmarking represent a considerable source of error.

Hands and feet measurements have made the focus of a distinct evaluation whose results are reported in table 7. Provided that those measurements are taken on relatively small body parts, it is not surprising to obtain lower mean errors than previously. The respective percent scores of the three studies are 25%, 25% and 40%.

Table 7 : MAD for Feet and Hands

Variable	AOE (mm)	USA (mm)	T-1 (mm)	T-2 (mm)	US%	T1%	T2%
095-Foot Length	3	0.5	0.1	0.8	17	3	27
094-Ball of Foot Length	6	0.8	1.1	1.6	13	18	27
096-Foot Breadth Horizontal	2	0.6	0.6	0.7	29	30	35
097-Bimalleolar Breadth	2	0.5	0.1	0.4	24	7	20
098-Heel Breadth	2	0.8	0.1	0.8	38	7	40
131-Hand Length	3	1.1	1.3	2.4	36	44	80
132-Hand Breadth	2	0.4	1.0	0.6	19	50	30
133-Hand Thickness	3	-----	0.8	0.7	-----	28	22
134-Thumb Breadth	2	0.2	0.0	0.1	10	0	5
130-Wrist-Index Length	4	0.9	1.2	2.1	22	29	53
129-Wrist-Thumbtip Length	3	1.0	1.0	2.0	34	33	68
128-Wrist-Center of Grip	4	1.5	2.0	2.8	38	50	69
Mean	**3**	**0.7**	**0.8**	**1.3**	**25**	**25**	**40**

AOE = Allowed Observer Error; USA = United States Army; US% = USA / AOE
T-1 = Test 1 (n=10) T1% = T-1 / AOE; T-2 = Test 2 (n=17); T2% = T-2 / AOE

The four least reliable measures, in that case, were hand length (80%), wrist-index length (53%), wrist-thumbtip length (68%) and wrist-centre of grip length (69%). All of these variables were measured with respect to one anatomical landmark: the stylion. Slight variations in landmarking from one time to another can account for the whole proportion of these errors.

Table 8 shows the results concerning head and face measurements. For all three studies, these results are excellent in absolute values (1.1, 0.8, 1.0) as well as in percent scores (28%, 23%, 24%). Most of these measurements are based on well-defined landmarks (eyes, nose, ears) or on bone features (head length, head breadth, bizygomatic breadth). Nevertheless, the cranio-facial arcs where both tragions have to be marked, have a higher proportion of errors in our test #2 than in the other studies. The observed errors are, however, always below the AOE.

Table 8 : MAD for Head and Face

Variable	AOE (mm)	USA (mm)	T-1 (mm)	T-2 (mm)	US%	T1%	T2%
139-Head Length	2	0.6	1.3	0.6	29	64	30
140-Head Breadth	2	0.5	0.4	0.3	25	21	12
141-Bizygomatic Breadth	2	0.6	0.3	0.3	30	14	15
142-Interpupillary Breadth	2	0.2	1.0	0.3	11	50	12
136-Menton-Sellion Height	3	1.0	0.7	0.8	34	24	27
137-Ear Length	2	0.5	0.4	0.3	25	21	15
138-Ear Length Above Tragion	2	0.6	0.3	0.5	31	14	23
143-Ear Breadth	3	0.7	0.4	0.9	24	14	30
144-Ear Protusion	3	0.7	0.6	1.3	25	19	41
148-Bitragion Submandibular Arc	6	2.7	1.3	2.6	45	21	42
149-Bitragion Chin Arc	8	1.4	0.9	1.6	17	11	19
150-Bitragion Subnasale Arc	6	1.5	0.1	1.1	24	2	18
151-Bitragion Frontal Arc	5	1.7	1.4	1.1	34	29	21
152-Bitragion Crinion Arc	5	1.8	0.7	1.1	36	14	21
153-Bitragion Coronal Arc	7	2.2	1.9	3.1	31	27	44
Mean	**3.9**	**1.1**	**0.8**	**1.0**	**28**	**23**	**24**

AOE = Allowed Observer Error; USA = United States Army; US% = USA / AOE
T-1 = Test 1 (n=10) T1% = T-1 / AOE; T-2 = Test 2 (n=17); T2% = T-2 / AOE

Table 9 shows the results for all sitting position measurements. The observed values are quite similar to those seen in previous tables: 48%, 37% and 43%. The highest MAD value in our test #2 is the sitting knee height. In this particular case, the observer had to measure the vertical distance between the platform where the feet lie and landmark above the knee, this point having been marked while the subject was standing. The shift of position combined with differences in landmarking accounts for a large part of this error.

Table 9 : MAD Sitting measurements

Variable	AOE (mm)	USA (mm)	T-1 (mm)	T-2 (mm)	US%	T1%	T2%
004-Sitting Height	6	3.0	1.6	2.9	49	27	48
154-Eye Height Sitting	8	4.1	2.3	1.6	51	28	19
155-Cervicale Height Sitting	10	2.7	2.4	4.7	27	24	47
156-Midshoulder Height Sitting	9	3.5	1.3	3.2	39	14	36
157-Acromial Height Sitting	9	4.4	1.8	2.1	49	19	23
162-Elbow Rest Height	10	4.7	5.9	2.2	47	59	22
160-Waist Height n.i. Sitting	6	2.8	2.9	4.1	46	48	68
161-Waist Height o. Sitting	8	3.2	2.4	2.7	40	30	33
163-Thigh Clearance	3	1.5	1.5	1.9	50	50	63
164-Knee Height Sitting	2	0.8	1.8	1.8	38	88	90
165-Popliteal Height	7	2.3	2.9	3.8	33	41	54
166-Wrist Height Sitting	10	7.9	6.1	2.8	79	61	28
167-Forearm-Forearm Breadth	17	8.5	2.6	4.3	50	15	25
168-Hip Breadth Sitting	6	2.4	1.6	3.7	40	27	61
169-Buttock-Knee Length	6	3.8	1.9	2.6	63	31	43
170-Buttock-Popliteal Length	7	4.7	3.5	2.4	67	50	34
171-Functional Leg Length	17	6.8	3.4	6.5	40	20	38
Mean	**8.3**	**3.9**	**2.7**	**3.1**	**48**	**37**	**43**

AOE = Allowed Observer Error; USA = United States Army; US% = USA / AOE
T-1 = Test 1 (n=10) T1% = T-1 / AOE; T-2 = Test 2 (n=17); T2% = T-2 / AOE

Weight and skinfolds are of particular interest (table 10). In the case of weight variable, errors from both U.S. Army and test #1 range between 100 and 200g. If one takes into account the daily weight variation of one subject, which is proportional to food intake and the elimination of waste (1dl = ±100g), these errors are not significant.

Though, in our test #2, the mean observed error is about 2kg which is far too high. But, if we keep in mind that the data have been gathered throughout a 2 month period, this variation is not surprising. Furthermore, one of our standard subjects was obese and varied more than the others (from 140 to 151.6 kg for a height of 1727 mm).

Skinfolds were not measured in the U.S. Army survey. Hence we have arbitrarily established a 5 mm threshold for all related AOEs on the basis of our own experience, given that these measures are amongst the least reliable. The average observed errors happen to be 1.4 mm and 1.3 mm in both our tests, the highest error value being for thigh skinfold (3.6 and 1.7 mm). This skinfold is rather difficult to measure, especially if the subject is characterized by a surplus of fatty tissues or by cellulite.

Table 10 : MAD for Weight (kg) and Skinfolds

Variable	AOE (mm)	USA (mm)	T-1 (mm)	T-2 (mm)	US%	T1%	T2%
001-Weight	0.3	0.1	0.2	2.0	33	53	666
195-Subscapular Skinfold	5	-----	1.5	1.4	-----	30	28
196-Abdomen Skinfold	5	-----	1.5	1.0	-----	30	20
197-Supra-iliac Skinfold	5	-----	2.0	0.9	-----	40	17
198-Thigh Skinfold	5	-----	3.6	1.7	-----	71	33
199-Biceps Skinfold	5	-----	0.5	1.0	-----	11	20
200-Triceps Skinfold	5	-----	0.7	1.1	-----	13	22
Mean	**4.3**	**0.1**	**1.4**	**1.3**	**33**	**35**	**115**

AOE = Allowed Observer Error; USA = United States Army; US% = USA / AOE
T-1 = Test 1 (n=10) T1% = T-1 / AOE; T-2 = Test 2 (n=17); T2% = T-2 / AOE

Finally, the reaches are listed in table 11. The mean error differences between the three tests are very sizable for that specific set of variables. For the 7 reaches directly taken, the U.S. Army's mean MAD was 10.7 mm. In our test #1, the mean MAD was only 3.4 mm. It is therefore highly probable that the differences between both studies is related to motivation, since the subject's performance (reaching the furthest possible point) is encouraged by the observer. However, in our test #2, the MADs are, on average, 12.3 mm, which is considerable. Moreover, the differences in landmarking cannot account for that error if we exclude the case of wrist-wall lengths. Perhaps can we invoke the motivation hypothesis once again, but this argument alone cannot explain the huge variation between the three studies.

Table 11 : MAD for Reaches

Variable	AOE (mm)	USA (mm)	T-1 (mm)	T-2 (mm)	US%	T1%	T2%
010-Span	10	7.3	3.2	9.0	73	32	90
172-Ov.Fing.Reach	20	12.1	5.3	6.3	60	27	31
173-Ov.Fing.Reach Ext.	20	10.9	2.0	8.7	54	10	43
189-Thumb.Reach	20	10.7	2.7	17.3	53	13	87
193-Wrist-Wall Length	20	11.0	2.7	14.1	55	13	70
194-Wrist-Wall Length Ext.	20	12.7	4.0	14.1	63	20	70
180-Ov.Fing.Reach Sitting	20	10.4	4.0	16.6	52	20	83
Mean	**18.6**	**10.7**	**3.4**	**12.3**	**59**	**19**	**68**

AOE = Allowed Observer Error; USA = United States Army; US% = USA / AOE
T-1 = Test 1 (n=10) T1% = T-1 / AOE; T-2 = Test 2 (n=17); T2% = T-2 / AOE

Conclusions

If we only consider the results of the 126 measurements from the first seven tables, we realize that the mean errors only represent small variations: 2.95 mm for the U.S. Army (43% of the AOE which is 6.94 mm), 1.89 mm for our test #1 (27% of the AOE) and 2.68 mm for our test #2 (38% of the AOE).

According to the actual norms, these errors can be considered as acceptable and show that the mean values gathered during our fieldwork among the Canadian Army Land Forces are as valid as those gathered in other surveys of the same kind.

However, it is noteworthy to point out that our results show that the total error does not solely originate from differences between two observers. There is also a non-negligible intraobserver error that exists independently of the landmarking process. Furthermore, our test #2 clearly shows that a considerable part of the anthropometric error of measurement is not due to either intraobserver or interobserver error, but originates from the subjects' landmarking as well as from intraindividual variations through time or from the changing motivation of the subjects to follow the observer's instructions.

Those errors, even if they appear «acceptable» from a technical standpoint, may have some repercussions if anthropometric data are used for ergonomic purposes (Gordon, 1992). Hence, one should always keep these errors in mind when drawing conclusions from statistical anthropometric data.

In fact, most of the specialists agree that the total variance is composed of many factors: the dispersion of the individuals around their mean, the biological modifications they experience through time and the sum of all technical errors (errors in landmarking, interobserver and intraobserver errors of measurement). Each of these errors contributes to thhe inflation of the biological variance. Even if one works to counteract this inflation, no biological data will ever be devoid of error. One can only attempt to keep it within reasonable limits, and this was our main purpose while collecting the Canadian Army data.

From the users' standpoint, this artificial inflation means that the published variances are conservative in that the real biological variances can only be lower than the computed values. Hence, estimated limits that should account for 95% of the targeted population would in fact account for a slightly higher proportion of the population. This is what one logically seeks. Therefore, these tables can be used with the confidence generally attributed to such results.

References

Dahlberg G., 1940: "Statistical Methods for Medical and Biological Students". London: Allen and Unwin.

Chamberland, A., Carrier, R., Forest, F., & Hachez, G. (1998). *Anthropometric survey of the Land Forces (LF97)* (Contractor report 98-CR-15). Toronto, Ontario: Defence and Civil Institute of Environmental Medicine.

Gordon C.C., Churchill T., Clauser C.E., Bradtmiller B., McConville J.T., Tebbetts I., Walker R.A., 1989: "*1988 Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics*". Technical Report NATICK/TR-89/044, United States Army Natick, Research, Development and Engineering Center, Natick, Massachusetts 01760-5000 : 638 pages.

Gordon C.C., Bradtmiller B., 1992: "*Interobserver Error in a Large Scale Anthropometric S*urvey". Am. J. Hum. Biol. 4: 253-263.

Russel C.M., Williamson D.F., Bartko J.J., Bradley E.L., 1994: "*Simulation Study of a Panel of Reliability Indicators Applied to Paired Measurements*". Am. J. Hum. Biol. 6: 311-320.

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM
(Highest classification of Title, Abstract, Keywords)

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g., Establishment sponsoring a contractor's report, or tasking agency, are entered in section 12.) Genicom Consultants	2. DOCUMENT SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable) Unclassified

3. DOCUMENT TITLE (the complete document title as indicated on the title page. Its classification should be indicated be the appropriate abbreviation (S,C,R or U) in parentheses after the title.)
Anthropometric survey of the Land Forces: measurement error analysis

4. DESCRIPTIVE NOTES (the category of the document, e.g., technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)
Contractor report

5. AUTHOR(S) (Last name, first name, middle initial. If military, show rank, e.g. Burns, Maj. Frank E.)
Forest, Francis, Chamberland, André, Billette, Jean-Michel, and Meunier, Pierre

6. DOCUMENT DATE (month and year of publication of document) January 1999	7.a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.) 16	7.b. NO. OF REFS. (total cited in document) 5

8.a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)	8.b. CONTRACT NO. (if appropriate, the applicable number under which the document was written) W7711-6-7350
9.a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)	9.b. OTHER DOCUMENT NO.(S) (any other numbers which may be assigned this document either by the originator or by the sponsor.) DCIEM CR 1999-041

10. DOCUMENT AVAILABILITY (any limitation on further dissemination of the document, other than those imposed by security classification)

x	Unlimited distribution
	Distribution limited to defence departments and defence contractors; further distribution only as approved
	Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
	Distribution limited to government departments and agencies; further distribution only as approved
	Distribution limited to defence departments; further distribution only as approved
	Other

11. ANNOUNCEMENT AVAILABILITY (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (10.) However, where further distribution (beyond the audience specified in 10) is possible, a wider announcement audience may be selected.)

12. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)
Human Engineering Sector
Defence and Civil Institute of Environmental Medicine
PO Box 2000, 1133 Sheppard Avenue West
North York, Ontario, M3M 3B9

DSIS DCD03
HFD 09/94

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM
(Highest classification of Title, Abstract, Keywords)

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

The Canadian Forces carried out an anthropometric survey for which the results were published in 1998. In order to evaluate the precision and accuracy of the measurements performed in the survey, an error analysis study was performed in parallel with the main survey. Precision, which is the difference between repeated measurements during the survey, was first analysed. . For this portion of the analysis, each of the last seven participants of the day were assigned to a particular station in order to be reassessed for a given set of variables after the whole session at the end of each day. Each subject had the same landmarks.

Accuracy, which is the difference between the measurements during the survey and the true value, was also monitored throughout the survey to avoid the emergence of systematic bias. To this end, 4 individuals from the survey's team of observers (the landmarkers) were selected. The "true value" was obtained through averaging the values obtained by a group of trained measurers for each of the anthropometric variables. Each of the 4 individuals was re-landmarked and re-measured at regular intervals throughout the survey, i.e. over a period of 2 months. This enabled the survey team to control the quality of the measurements by addressing landmarking and measurement biases.

The comparison of these two data sets with the US Army's MADs (Mean Absolute Difference between two measurements) and AOEs (Allowable Observer Error) was performed for the 140 survey measurements. The results show the measurement errors of the Land Forces survey to be comparable in magnitude to those obtained in the 1988 US Army survey.

This study highlights the occurrence of errors in measurements examined twice by the same examiner, errors resulting of the landmarking as well as variations during the time frame (up to two months) for the same measurement of the same subject.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible, keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Anthropometric survey, error analysis

#511090