Prosthetics and Orthotics International, 1999, 23, 260–273

Comparative trials on hybrid walking systems for people with paraplegia: an analysis of study methodology

M. J. IJZERMAN*/**, G. BAARDMAN*, H. J. HERMENS*, P. H. VELTINK**, H. B. K. BOOM** and G. ZILVOLD*

**Roessingh Research and Development b.v., Enschede, The Netherlands*
***University of Twente, Institute for Biomedical Technology, Enschede, The Netherlands*

Abstract

A new orthosis (SEPRIX) which combines user friendliness with low energy cost of walking has been developed and will be subject to a clinical comparison with conventional hip-knee-ankle-foot orthoses. In designing such comparative trials it was considered it may be worthwhile to use previous clinical studies as practical examples. A literature search was conducted in order to select all comparative trials which have studied two walking systems (hip-knee-ankle-foot orthoses) for patients with a complete thoracic lesion. Study population, intervention, study design, outcome measurement and statistical analyses were examined. Statistical power was calculated where possible.

Of 12 selected studies, 7 were simple A-B comparisons, 2 A-B comparisons with a replication, 2 cross-over trials and 1 non-randomised parallel group design, the last of which was considered internally invalid due to severe confounding by indication. All A-B comparisons were considered internally invalid as well, since they have not taken into account that a comparison of two orthoses requires a control for aspecific effects (like test effects) which may cause a difference. Statistical power could only be examined in 4 studies and the highest statistical power achieved in one study was 47 %. It is concluded that statistical power was too low to be able to detect differences. Even analysis through interval estimation showed that the estimation of the difference was too imprecise to be useful. Since the majority of the surveyed papers have reported small studies (of only 4-6 patients), it is assumed that lack of statistical power is a more general problem. Three possibilities are discussed in order to enhance statistical power in comparative trials, i.e. multicentre studies, statistical pooling of results and improving the efficiency of study design by means of interrupted time series designs.

All correspondence to be addressed to Maarten J. IJzerman PhD, Roessingh Research and Development, Roessinghsbleekweg 33b, 7522 AH, Enschede, The Netherlands. Tel: (+31) 53 4875777; Fax: (+31) 53 4340849; e-mail: m.ijzerman@rrd.nl

Introduction

Various new developments in hybrid systems for patients with paraplegia have been reported in the last few years (Stallard and Major, 1995; Edwards and Bataweel, 1996; Ferrarin and Rabuffetti, 1996; Yang *et al.*, 1996). Beside technological improvements, two categories of clinical research can broadly be distinguished, i.e. comparisons between available clinical systems (like Reciprocating Gait Orthoses - RGO) and clinical trials on new system developments. Reduction in energy cost of locomotion is often a main goal in these reports. However, in the design of a new orthosis (Separable Reciprocator with Intelligent Knee Stabilisation - SEPRIX) the authors acknowledged user friendliness, in addition to energy cost of locomotion, to be an important target. Separability and foldabability were suggested in order to improve donning / doffing and transportation respectively (Baardman *et al.*, 1997), whereas upper body load during locomotion can be reduced by aligning an

orthosis in slight abduction (IJzerman *et al.*, 1997[b]).

The effectiveness of a new walking system should be investigated in order to establish superiority of the system over another. In claiming such effectiveness, in particular when the decisions are complicated (like justification of implant technology) and the differences are small, appropriate methodology is essential (Yusuf *et al.*, 1994; Sykes *et al.*, 1996).

In designing clinical studies on effectiveness of orthoses, previous studies are often used as examples. For instance, the study of Sykes *et al.* (1996) is, with respect to the methodology, almost similar to the study of Nene and Patrick (1990). However, in designing a study on effectiveness of the new SEPRIX orthosis, some methodological problems were encountered in previously published studies and copying the methodology appears not to be a suitable approach.

The basic issues in the design stage of an experiment are usually categorised into internal validity and external validity. Internal refers to the validity of the inferences drawn in regard to the patients involved in the study and external refers to the validity of the inference in regard to the target population outside the sample (Rothman, 1986). Internal validity is a prerequisite for external validity, but external validity is not always likely to be a consequence of an internally valid study. In addition, statistical conclusion validity concerns whether the differences in the study are due to chance (Cook and Campbell, 1979; Wagenaar, 1990; Ottenbacher and Barrett, 1990). In a statistically valid study one presumes that appropriate statistical tests were used and that the study had sufficient statistical power to detect differences (Ottenbacher and Barrett, 1990).

In general, true-experimental designs are considered the most powerful method in showing treatment effectiveness (Ottenbacher, 1995). Although other designs can also be considered as true experimental (Cook and Campbell, 1979; Wagenaar, 1990), the randomised controlled trial (RCT) is uniformly accepted as the most appropriate design (Meinert, 1986; Pocock, 1983; Reilly and Findley, 1989; Pollock *et al.*, 1993). Control in this context refers to a comparison of the new therapy with a standard therapy, randomisation refers to random assignment of patients to either a new or a standard therapy (Pocock, 1983). However, RCTs are difficult to conduct (Reilly and Findley, 1989) and, so far, no comparison of walking systems has been a randomised controlled trial. Two reasons can be put forward, i.e. (1) adequate randomisation usually requires sufficient patients in order to achieve baseline comparability of study groups (potential confounders) (Rothman, 1986; Pollock *et al.*, 1993) and (2) if only a limited number of patients can be included, an RCT is not the most efficient design, since 50 % of the patients are assigned to a standard orthosis. In contrast, cross-over trials do offer true experimentation but with a maximum statistical power since all patients are assigned to both standard as well as new therapy (Pocock, 1983; Senn, 1993). However, cross-over trials have the disadvantage that a carry-over effect may limit their internal validity.

With respect to the statistical analysis of studies, it is often found that much attention is given to proper statistical testing, but that statistical power is not addressed (Ottenbacher and Barrett, 1990). In addition, the majority of the comparative trials on paraplegic walking systems were not able to demonstrate significant differences between walking devices. Proper interpretation of the p-value (accepting H_0 when $p>\alpha$ and rejecting H_0 if $p<\alpha$) would then suggest that H_0 (no difference) is true (Barnard, 1990; Salsburg, 1990). However, this interpretation neglects the risk of a type II error (β-level). A type II error occurs if H_0 is accepted falsely or in other words there truly is a difference but the study fails to show it because of lack of statistical power (Altman, 1980; Lachin, 1981; Carpenter, 1993; Ottenbacher, 1995). Studies with lack of statistical power are unable to detect any clinically relevant difference or they provide estimates which are too imprecise to be useful (Carpenter, 1993). Conduct of such trials with *a priori* knowledge of insufficient statistical power is considered unethical, since patients and other resources are used without having a reasonable chance of drawing correct conclusions (Altman, 1980).

This study aims to support the design of clinical trials on effectiveness of (hybrid) walking systems for patients with paraplegia. A database search was carried out and the relevant literature is summarised with respect to study sample, intervention (walking system and gait

training), study design, main outcome measures and statistical analysis. The methodology will be discussed with respect to the internal validity, statistical conclusion validity and external validity. Finally, some alternative approaches for clinical evaluations of walking systems are discussed.

Methods

Literature search

A database search was performed in order to obtain all relevant publications concerning a clinical evaluation of a walking system for patients with (complete) paraplegia. Publications were searched in the Medline database in the period from January 1966 to May 1997. Search key-words included paraplegia, orthosis, walking, energy, electrical stimulation either solely or in various combinations. Current contents as well as references listed in available publications were scanned, since it is possible that a database only partially yields the available publications (Sacks *et al.*, 1987).

Publications were included in the analyses if they had met the following conditions: (1) the comparison should be performed in adult patients suffering from complete thoracic paraplegia, (2) the comparison should comprise hip-knee-ankle-foot orthoses, either with or without electrical stimulation, (3) the trial should report a clinical comparison of two orthoses and (4) the trial should have used relevant clinical outcome measures like oxygen uptake, walking speed or crutch forces.

Summary of relevant study aspects of the selected studies

All selected studies were summarised regarding 5 different aspects, i.e. study sample, type of walking system which is compared, actual study design, outcome assessment and statistical analysis (Table 1).

Table 1. Relevant aspects along which the selected articles are judged. The five aspects include study population, intervention, study design, outcome measures and statistical analysis. Each aspect was judged on one or more criteria which are listed in third column of the table. If applicable, abbreviations to be used in Table 2, are listed in the last column. *Abbreviations:* RGO: reciprocating gait orthosis, HGO: hip guidance orthosis (parawalker), ARGO: advanced reciprocating gait orthosis, iRGO: isocentric Reciprocating Gait Orthosis, FES: Functional Electrical Stimulation, Quads: quadriceps muscles, Hams: Hamstrings, Glut: Gluteal muscles, V_{02}: oxygen uptake (ml/min), E_{02}: oxygen cost (ml/m), v: walking speed, CFTI: crutch time force intergral, CPF: crutch peak force, SD: standard deviation.

Study aspects	Ctiterion	Abbreviations
Study sample	A. Description of sample	
	• size	–
	• gender	m: male ; f: female
	• age	–
	• level	–
	B. Previous walking experience	① yes : ② no
Intervention	C. Standard system	
	D. New system	
	• orthosis	① RGO : ② HGO : ③ ARGO : ④ iRGO
	• FES application	a. Quads/hams ; b. Glut/hams ; c. Glut
	E. Adequate training period	① gait training described ② prior to measurements
Study design	F. Design type	① within group (AB) ② within group (AB/BA cross-over) ③ between group
	G. Measurement day	① same occasion ② different occasion
	H. Measurement sequence	① randomised ② fixed
Outcome measure	I. Main outcome measures	① V_{02} ; ② E_{02} ; ③ v ; ④ CFTI ; ⑤ CPF
	J. Assessment speed	① self-selected ; ② standardised
Statistical analysis	K. Descriptive statistics	① mean ; ② SD a. of differences ; b. for each system
	L. Inferential statistics	① parametric/non-parametric tests ② analysis of variance

With respect to the *study sample* the relevant baseline characteristics of the included subjects were of particular interest. Articles were judged on the actual sample size, level of lesion, gender, age and walking experience of the subjects included in the study.

The *walking system* being evaluated in the selected studies could be either a standard (conventional) orthosis or a new system. Four conventional orthoses are investigated frequently, i.e. the Reciprocating Gait Orthosis (RGO), Hip Guidance Orthosis (HGO), Advanced Reciprocating Gait Orthosis (ARGO) and isocentric® Reciprocating Gait Orthosis (iRGO). Three different applications of Functional Electrical Stimulation (FES) are used, i.e. (1) reciprocal electrical stimulation of quadriceps/hamstrings, (2) electrical stimulation of hamstrings / gluteal muscles on stance side and (3) electrical stimulation of gluteal muscles on stance side.

Judgement of the contents and execution of the gait training is difficult, but was considered adequate if the authors had described an extensive and adequate training schedule for the orthosis gait training, FES muscle conditioning as well as hybrid system gait training.

Judging the actual *study design* as being either experimental (randomised parallel group design) or quasi-experimental (non-randomised parallel group design) is not adequate as most authors have conducted a within patient comparison (each subject is tested in either of the two walking systems). However, within subject comparisons can be carried out differently. Methodologically most profound is a study which is conducted as a cross-over design. Randomisation of the measurement sequence (either A-B or B-A, which essentially provides a cross-over design) is considered an attempt to improve the internal validity of the A-B design. A simple A-B comparison without randomisation, i.e. all subjects are tested in the same order, is least favourable with respect to internal validity.

Measurement of the performance in either orthosis (*outcome assessment*) could have been conducted on either one or on separate days (each orthosis on a different day). This is important as the performance in the second orthosis if measured on one day may be affected by either fatigue of the subjects or decreased FES muscle performance. Though measurement of the performance of two walking systems on different days may be superior with respect to fatigue, they are more affected by day to day variation in performance.

Important *outcome measures* are those which convey information to be used for decision making with respect to the performance of walking. Though arbitrary, important outcome measures are: oxygen uptake (oxygen uptake: V_{O2}, oxygen cost: E_{O2}), walking speed (v), or crutch force measures (crutch force time integral: CFTI or crutch peak force: CPF). Assessment of the walking performance could have been performed at either a standardised or a self-selected walking speed.

At least descriptive *statistics* should have been reported by the authors. The central tendency and the variation of each walking system should have been described as well as the differences between the walking systems. Statistical testing in order to draw inferences was judged with respect to the appropriateness of the tests (dependent/independent group comparisons and parametric/non-parametric) as well as the required statistical assumptions.

Post hoc statistical power calculations

All data with respect to V_{O2} and walking speed for each orthosis in each article are summarised including the p-level reported for the difference between the systems. *Post hoc* statistical power calculations were performed using the statistical program PC-size (Dallal, 1990). As most studies were within-subject comparisons (A-B), statistical power was calculated assuming that the analysis was conducted using a paired t-test. Since this test requires the mean and standard deviation of the differences between orthoses, statistical power could only be calculated if these data were available.

Results

Selected studies

Most studies were identified while using the key-words "orthosis" and "paraplegia" separately. No search strategy could be determined to locate studies on hybrid walking systems for patients with paraplegia specifically. In addition to the papers which were excluded according to the selection criteria, papers were also excluded if they reported a case study (Phillips, 1989; Isakov *et al.*, 1992; Jefferson and Whittle, 1990; Muszkat *et al.*, 1994; Smith

et al., 1997) or if they presented results of a single system with reference to the literature (Gallien *et al.*, 1995; Saitoh *et al.*, 1996; Nene and Patrick, 1989; Nene and Jennings, 1992). The latter articles were excluded because such comparisons are often distorted by many other effects which are more related to subject selection and assessment protocols than to the differences between orthoses. A paper of Stallard and Major (1995) was excluded since a later version of the HGO with stiff hip joints was compared with historical controls, which is also considered inadequate (Pocock, 1983). If two or more publications were found which reported from the same study, only the most recent publication was included. However, two publications from the ARMOR association were included because they reported with a different scope (Thoumie *et al.*, 1995; Beillot *et al.*, 1996). After inclusion, twelve papers were available for evaluation.

Summary of relevant study aspects of the selected studies

All details with respect to study design of the selected papers are given in Table 2.

The papers showed considerable concordance in the size and characteristics of the study population. Between 4 and 6 patients were included in the majority of the studies. Two larger studies comprised 28 and 22 patients respectively (Lotta *et al.*, 1994; Whittle and Cochrane, 1989). Mean age ranges from approximately 23 to 38. The majority of the patients were male. All studies included patients with low as well as high thoracic lesions. Two studies also reported results of patients with incomplete lesions (Sykes *et al.*, 1996; Winchester *et al.*, 1993). There were two common approaches for inclusion of patients. Either new patients were recruited and trained with two new systems (5 studies) or patients with previous walking experience were trained with a new system (6 studies).

Walking systems which are compared in the literature comprise very often the RGO. Only a few papers have been published on HGO, ARGO and iRGO.

Most authors reported extensively on the training schedules used in their study, including the stimulation equipment and parameters settings. Gait training was conducted prior to all measurements in all but one study (Whittle and Cochrane, 1989). They conducted a cross-over trial and the gait training in the second orthosis was performed after the subject had been

Table 2. Summary of the relevant study aspects of the included studies (see Table 1 for explanation).

	Study sample					Intervention			Study design			Outcome measures		Statistical analysis	
	A				B	C	D	E	F	G	H	I	J	K	L
Paper	size	gender	age	level											
Hirokawa *et al.*, 1990	6	4,2	34.7 (8.2)	T1-T10	2	1	1a	1,2	1	?	?	1,2	2	2a	–
Sykes *et al.*, 1996	5	4,1	31.8 (5.3)	C2-T9	1	1	1a	1,2	1	1	2	1,2,3	1	1a,2a[4]	–
Petrofsky and Smith, 1991	4	?	20-35	T4-T12	3	1	1b	1,2	1	?	?	1,2,5	2	1b,2b	?
Thoumi *et al.*, 1995	6	6,0	35	T2-T10	2	1	1a	2	1	?	?	3	1	1b,2b	2
Beillot *et al.*, 1996	4	4,0	31	T7-T11	2	1	1a	2	1	1	1	1,2	2	1a,1b	1
Winchester *et al.*, 1993	4	4,0	29.8(6.1)	T5-T10[1]	1	1	4	1,2	1[3]	2	1	1,2,3	1	1b,2b	2
Whittle and Cochrane, 1989	22	18,4	33.6(5.8)	T3-T12	2	1	2		2	2	–	1,2,3,4[6]	1	1a,1b	–
Lotta *et al.*, 1994	4	4,0	23.8(3.8)	T3-T12	2	2	–	1,2[5]	3	–	–	3	1	1a,2a[4]	1
	11	10,1	23.5(5.8)	T3-T10	2	3	–	1,2[5]		–	–	3	1		
	13	10,3	25.3(8.6)	T3-T12	2	1	–	1,2[5]		–	–	3	1		
McClelland *et al.*, 1987	3	?	?	T4-T7	1	2	2c	1,2	1	1	?	3,4	1	1a,2a[4]	–
Nene and Patrick, 1990	5	5,0	27.6(1.8)	T4-T7	1	2	2c	1,2	1	1	2	1,2,3	1	1a,2a[4]	–
IJzeman *et al.*, 1997[a]	6	6,0	38.7(11)	T4-T12	1	3	3[2]	1,2	1[3]	2	2	1,2,3,4,5	1	1a,2a[4]	1
IJzeman *et al.*, 1997[b]	5	4,1	36.8(6.0)	T4-T12	1	3	3[2]	1,2	2	2	–	1,2,3,4,5	1	1a,2a[4]	1,2

[1] two incomplete paraplegic patients, [2]modified ARGO, [3]replicated before-after trial, [4]mean and SD available from original publication, [5]unequal training among participating centres, [6]crutch peak force estimated from ground reaction force.

assessed in the first orthosis.

The majority of the studies lack detailed information on the *study design*. Out of twelve studies, only one had performed a (non-randomised) parallel group design (Lotta *et al.*, 1994). Eleven studies reported within subject comparisons. Six studies reported simple A-B comparisons without randomisation of the measurement order (Nene and Patrick, 1990; McClelland *et al.*, 1987; Sykes *et al.*, 1996; Hirokawa *et al.*, 1990; Petrofsky and Smith, 1991; Thoumie *et al.*, 1995). Only Beillot *et al.* (1996) performed an A-B comparison with randomisation of the measurement order. Two studies reported simple A-B comparisons with a replication (Winchester *et al.*, 1993; IJzerman *et al.*, 1997[a]). Finally, two studies reported a cross-over design (Whittle and Cochrane, 1989; IJzerman *et al.*, 1997[b]).

Four studies, which used a simple A-B comparison, have reported that all measurements took place on the same day (Beillot *et al.*, 1996; Nene and Patrick, 1990; McClelland *et al.*, 1987; Sykes *et al.*, 1996). In three of them, the (conventional) orthosis without FES was always measured as the first system.

All authors included a resting period in between the measurements. Three simple A-B comparisons did not provide information on the measurement sequence and whether assessments were performed on different occasions (Hirokawa *et al.*, 1990; Petrofsky and Smith, 1991; Thoumie *et al.*, 1995).

Two authors conducted a study with a replication of the simple A-B comparison (Winchester *et al.*, 1993; IJzerman *et al.*, 1997[a]). Winchester *et al.* (1993) measured both orthoses twice on 2 separate days and has randomised the measurement order. IJzerman *et al.* (1997[a]) has measured both orthoses twice on four separate days according to a BABA sequence.

Oxygen uptake (V_{O2} : $ml.min^{-1}$) and oxygen cost (E_{O2} : $ml.m^{-1}$) were reported as *outcome measures* in eight studies. Self-selected walking speed is presented by all but three authors. Crutch forces were measured in four studies. Conversion of oxygen uptake ($ml.min^{-1}.kg^{-1}$) to energy uptake equivalents ($J.min^{-1}.kg^{-1}$) was performed by four authors (Hirokawa *et al.*, 1990; Nene and Patrick,1990; Beillot *et al.*, 1996; Sykes *et al.*, 1996)

Three studies performed the measurements at a standardised walking speed imposed by means of a metronome (Hirokawa *et al.*, 1990; Petrofsky and Smith, 1991; Beillot *et al.*, 1996). All other studies have measured walking performance at a self-selected speed.

The majority of the authors have used some *statistical test* in order to conclude whether differences can be considered significant. Most tests include parametric or non-parametric paired tests and the results are reported in terms of p-values. Two publications have used confidence intervals (IJzerman *et al.*, 1997[a]; IJzerman *et al.*, 1997[b]). None of the authors has determined the actual statistical power of their study if a test failed to show significance.

Table 3 presents the actual data and the differences between the orthoses which were compared in the selected study including the results of the statistical tests.

Post hoc statistical power analysis was performed if the study revealed non significant results. *Post hoc* statistical power calculations in a within subject comparison (A-B) can only be calculated accurately if the mean and standard deviation of the differences are available. Since only four authors presented a standard deviation of the differences, statistical power analysis could only be conducted for those particular studies. *Post hoc* statistical power analysis showed that statistical power was only between 9% and 47% (Table 3).

Based on the actual data for these 4 studies in Table 3 it is possible to calculate a statistical power curve in which the statistical power can be estimated if the sample size is increased. Figure 1a and Figure 1b present the statistical power curve with (theoretical) increase in statistical power as a function of the sample size while using the data from the 4 studies in Table 3. Assuming that a statistical power of approximately 80% is acceptable, it can be concluded that the minimum number of subjects to be included in the comparative trial of Sykes *et al.* (1996) is 10.

Discussion

Internal validity, statistical conclusion validity and external validity are considered to be the important aspects of study design. In the remaining part of this paper, each of these items will be discussed with reference to the literature which is surveyed in the previous section.

Table 3. Summary of the results for main outcome measures oxygen uptake (V_{02}) and walking speed (v). Actual statistical power is reported in the last column if mean difference (δ) and standard deviation (σ) of the differences were available.

	Standard orthosis	New orthosis	Difference in V_{02} (δ(σ)/p-value)	Power	Standard orthosis	New orthosis	Difference in v (δ(σ)/p-value)	Power
Hirokawa *et al.*, 1990			16% (?)/?		–	–	–	
Sykes *et al.*, 1996	2.32 (0.83)	2.57 (0.67)	-0.25 (0.24) / p<0.08[2]	43%	0.225(0.10)	0.247 (0.12)	0.022 (0.02) / p<0.07[2]	47%
Petrofsky and Smith, 1991			? (?) / p<0.01		–	–	–	
Thoumi *et al.*, 1995					0.21 (0.02)	0.20 (0.02)	0.01 (?) / NS	
Beillot *et al.*, 1996	0.73 (0.16)	0.86 (0.16)	-0.13 (?) / NS		–	–	–	
Winchester *et al.*, 1993	14.2 (1.8)	13.0 (1.4)	1.2 (?) / NS		0.211 (0.03)	0.225 (0.04)	-0.014 (?) / NS	
Whittle and Cochrane, 1989	618 (?)	593 (?)	25 (?) / NS		0.24 (?)	0.23 (?)	0.01 (?) / NS	
Lotta *et al.*, 1994	–	–	–		1. RGO: 0.243 (0.09)		1-2: p=0.029	
					2. HGO: 0.119(0.14)		2-3: p=0.022	
					3. ARGO: 0.186 (0.1)			
McClelland *et al.*, 1987	–	–	–		15.5 (4.24)	18.1 (4.49)	-2.55 (0.91) / p<0.55[2]	
Nene and Patrick, 1990	2.59 (0.25)	2.50 (0.35)	0.09 (0.17) / p<0.31[2]	15%	0.233 (0.03)	0.24 (0.04)	-0.007 (0.02) / p<0.53[2]	9%
IJzerman *et al.*, 1997[a]	18.0 (3.2)	17.2 (3.1)	0.8 (1.1) / [-0.36, 1.96][1]	31%	0.24 (0.11)	0.23 (0.13)	0.01 (0.03) / [-0.028,0.045][1]	9%
IJzerman *et al.*, 1997[b]	16.9 (3.6)	17.9 (3.2)	-1.0 (2.51) / [-4.1, 2.14][1]	11%	0.246 (0.11)	0.274 (0.1)	-0.028 (0.03) / [-0.06,0.005][1]	36%

[1] Confidence intervals as presented in IJzerman *et al.*, 1997[a] and 1997[b] were trasformed to absolute data rather than the relative difference with respect to ARGO. [2] Significance calculated using presented data and paired t-test.

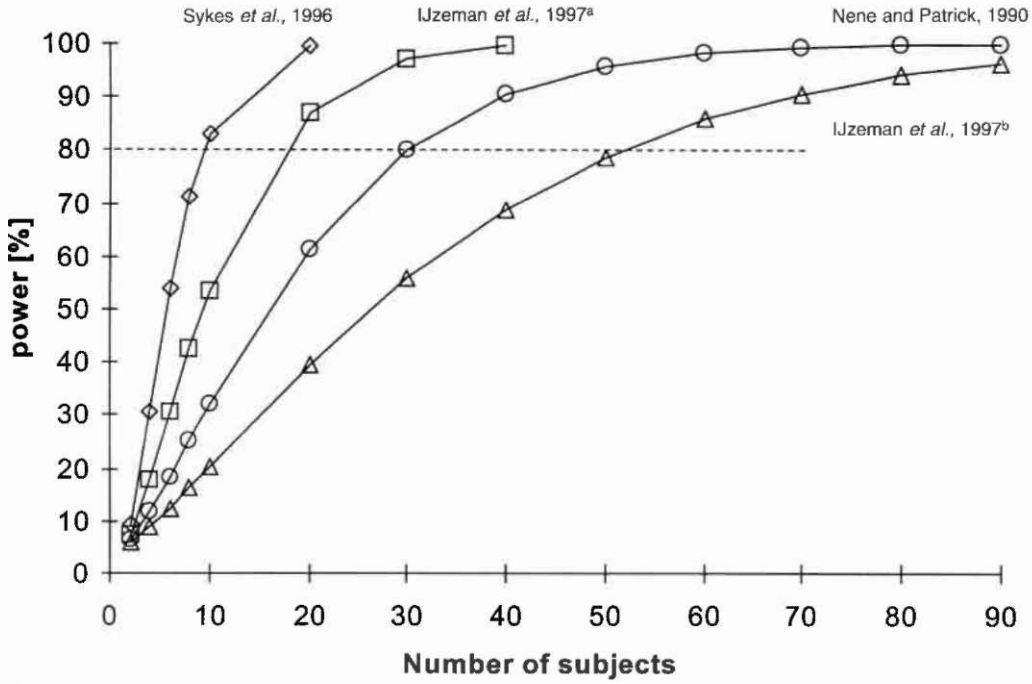

Fig. 1. (a) Statistical power of four studies (see Table 4 for actual statistical power) to detect a difference in V_{02} between two walking systems. The statistical power curve represents the theoretical increase in statistical power as a function of the sample size. The statistical power is calculated with the data of the indicated study and does not provide insight into the statistical power of a new trial. The data of Nene and Patrick (1990) and Sykes *et al.* (1996) were expressed in J. $min^{-1}.kg^{-1}$. These data were converted using the correction factor as presented in their articles.

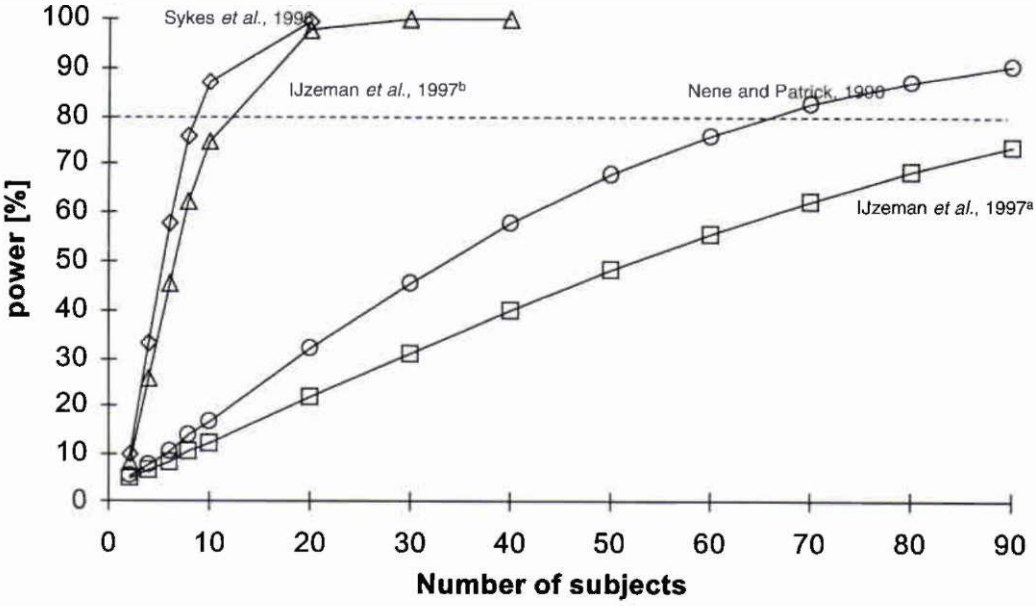

Fig. 1. (b) Statistical power curves calculated using the data of walking speed in four comparative studies. See for further explanation the text and the legends of Figure 1a.

Internal validity

Eleven (11) out of 12 studies were within patient comparisons, and one study was a between patient comparison. One major issue in judging the internal validity of parallel group trials is the baseline comparability of study groups (Meinert, 1986; Feinstein, 1985). Using randomisation an attempt is made in order to obtain baseline comparability with respect to relevant prognostic factors. However, Lotta *et al.* (1994) have not conducted a randomised controlled trial and baseline incomparability is present in their study. They allowed each of the participating centres to define their own selection criteria for each of the orthoses and subsequently, each centre executed their own training. As a result, HGO walkers (selected in only one centre) underwent less gait training than other patients in competing systems. Although Lotta *et al.* (1994) have noted that their data were not reliable, it must additionally be concluded that this comparative trial between HGO, RGO and ARGO to be internally invalid.

In judging the internal validity of a clinical trial, the trial should have acknowledged that there are specific and aspecific parts of a treatment, i.e. difference between orthoses. A control group provides a means with which the aspecific parts of the difference can be assessed. Any clinical trial which is conducted without such a control group is considered internally invalid beforehand, as the actual difference in the study between orthoses may be caused by various aspecific factors.

Aspecific effects which may be present in an A-B comparison of two orthoses are measurement errors, test-effects, history effects and regression to the mean (Cook and Campbell, 1979). *Test effects* comprise all differences which are caused by repeated use of the same testing procedures (Cook and Campbell, 1979). *History effects* are aspecific effects, caused in the period preceding the second measurement. For instance, if training in a new orthosis is conducted in between measurements, the difference can be partly explained by the level of experience with walking in general.

In a repeated measurement, a coincidental extreme (high or low) value in the first ("A") measurement is likely to be followed ("B" assessment) by a value closer to the mean. The magnitude of this *regression to the mean* depends on the reproducibility of the test-retest differences (Feinstein, 1985; Cook and Campbell, 1979).

Since the seven publications which have used a simple before-after comparison did not consider the need to control for these aspecific effects which may cause a difference, they lack internal validity. The observed differences comprise specific as well as aspecific parts.

Randomisation of the measurement sequence as performed by Winchester *et al.* (1993) and Beillot *et al.* (1996) might be a solution to average out the aspecific effects. However, if randomisation of the phase order is proposed, it is more appropriate to conduct a cross-over trial which essentially offers all possibilities to control and adjust systematic differences (period-effects) and thus enhances internal validity (Senn, 1993; Pocock, 1983).

Statistical conclusion validity

Statistical power

Only two studies have reported statistically significant results. Petrofsky and Smith (1991) reported a significant decrease in V_{O2} (normalised for walking distance) during the 1 mile walking test ($p<0.01$) but, except for a graph, they have not presented an average effect size. Lotta *et al.* (1994) presented significant differences between RGO - HGO and ARGO - HGO, but it is argued in a previous section that this study was internally invalid due to confounding by indication. Further analysis of the data of McClelland *et al.* (1987) showed a significant increase in walking speed in the hybrid Hip Guidance Orthosis.

Interpretation of the p-values of the other studies included in this survey implies that there is no statistical evidence of a difference. Statistical power ($1 - \beta$) could only be calculated in 4 "negative" trials and appeared to be between 10 and 50%. Although actual statistical power could not be calculated for the other trials, it is expected that, with the small differences in those studies, the statistical power will not exceed 50%. Assuming that statistical power should be between 80 and 90% (Lachin, 1981; Dupont and Plummer, 1990; Carpenter, 1993), it is evident that all comparative trials lack statistical power and have an unacceptably high risk of type II errors.

Interval estimation

A number of authors have acknowledged that p-values are often erroneously interpreted as evidence of a difference, but that they do not

give insight into the effect size which may be clinically relevant. (Altman and Bland, 1995; Carpenter, 1993; Barnard, 1990; Freeman, 1993; Rothman, 1986; Bulpitt, 1987). Instead, estimation of effect sizes by means of confidence intervals are favoured since they convey a summary of the original data in original units of measurement (Gardner and Altman, 1986; Rothman, 1986; Borenstein; 1994; Smith and Bates, 1992; Bulpitt, 1987).

Interval estimation has been recommended to enhance the interpretation of statistically non-significant trials in particular (Smith and Bates, 1992; Borenstein, 1994). Whereas calculation of statistical power in such "negative" trials can sometimes be misleading, a confidence interval conveys more information required to conclude whether there was sufficient statistical power to detect clinically relevant differences (Smith and Bates, 1992).

Figures 2a and 2b display 95% confidence intervals for the difference in V_{O2} and walking speed between two walking systems (see Table 3 for the actual data). A proper interpretation of this 95% CI is that out of 100 replicated studies, 95 will find a mean difference within this interval (Bulpitt, 1987).

The 95% CI for the mean difference in V_{O2} in the study of IJzerman *et al.* (1997[a]) suggests that V_{O2} is higher in the orthosis without reciprocal cable (95% CI [-2, 11%]). Though this study had insufficient statistical power (31%), it is concluded that this confidence interval provides sufficient information in order to conclude that no clinically relevant differences in favour of the orthosis without reciprocal cable can be expected. This clinically relevant difference was set at 20% in this study, since removing such reciprocal cable linkage results in a highly uncomfortable standing posture.

Sykes *et al.* (1996) found an increased walking speed in the RGO+FES of 10% on average (95% CI [-21%, 1%]). Replication of this study in a different group of patients may result in either a clinically relevant (-21%) or a difference which is negligible (1%). The same interpretation holds for the confidence intervals in most other studies. In these studies, more precision is required in order to conclude meaningful differences.

External validity

Two aspects are relevant in judging the external validity of a study, i.e. the target population of which the study population is selected and the selection criteria which are applied in order to obtain the actual study population.

In clinical studies on walking systems for people with paraplegia, the target population is considered to be "all people with paraplegia who may be appointed for prescription of a walking system".

Five studies have included patients who had no previous walking experience (Hirokawa *et al.*, 1990; Thoumie *et al.*, 1995; Beillot *et al.*, 1996; Whittle and Cochrane, 1989; Lotta *et al.*, 1994) and the other six studies (Sykes *et al.*, 1996; Winchester *et al.*,1993; McClelland *et al.*, 1987; Nene and Patrick, 1990; IJzerman *et al.*, 1997[a]; IJzerman et al, 1997[b]) have included subjects who already had experience with walking. Whereas the study population of the first five studies may be a representative sample of the previously defined target population, the study population of the latter six studies is in no case a representative sample. People who already have experience with walking and agree to participate in a new clinical trial are very motivated and do not represent the target population. As a consequence, external validity of at least the latter six studies is limited.

In order to obtain an internally valid study with sufficient precision, the study population is often restricted to a homogeneous population which is most likely to benefit from the orthosis. It is often not clear which selection criteria are applied to the study population. This may be explained by the fact that there are very few subjects available and that the use of narrow selection criteria will diminish the actual study population to zero.

For instance, all surveyed studies have included patients with low (T10-T12) and high thoracic lesions (T1 - T4) and none of the studies has used level of lesion as an inclusion criterion. This means that the study sample of the studies is heterogeneous. IJzerman *et al.* (1997[a]) reported a study in which level of lesion caused heterogeneity of effects. Two patients with low lesions had considerably lower walking speeds in the orthosis without reciprocating cable linkage, whereas four patients with T9-T12 lesions walked faster in that orthosis. Such heterogeneity in effect results in low statistical power if group analysis is performed. However, sub-group analysis is often not feasible, as the limited number of patients per sub-group

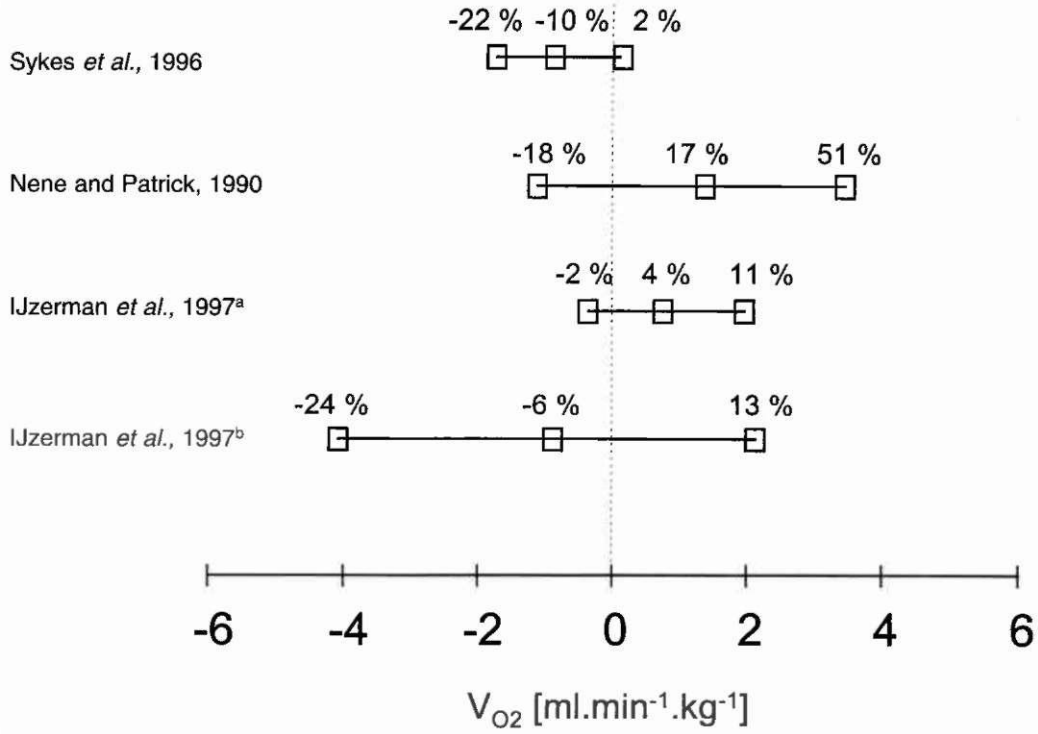

Fig. 2. (a) Estimation of difference in V_{02} between two walking systems by means of 95% confidence intervals. The confidence intervals are calculated with the data as given in the indicated study. Confidence interval in expressed as absolute difference (x-axis). Relative differences with respect to the standard orthosis are presented on top of each interval. The data of Nene and Patrick (1990) and Sykes *et al.* (1996) were expressed in $J.min^{-1}kg^{-1}$. These data were converted using the correction factor as presented in their articles.

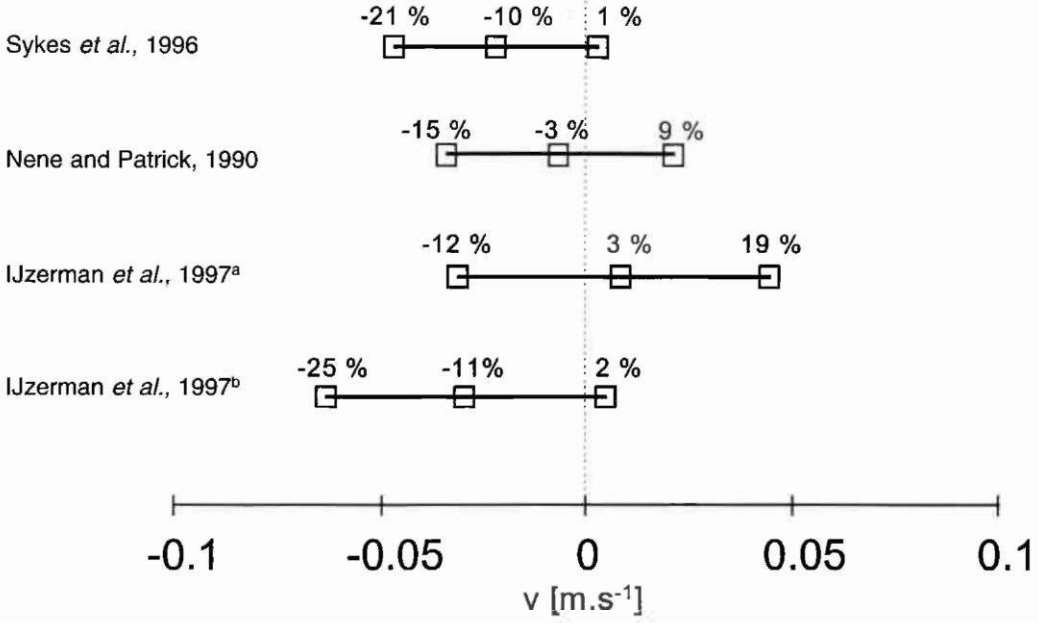

Fig. 2. (b) Estimation of difference in walking speed between two walking systems by means of 95% confidence intervals. See for further explanation the text and the caption of Figure 2a.

reduces statistical power. In general, restriction of the study population to a homogeneous group will enhance study efficiency, provided that sufficient patients can be included (Kleinbaum *et al.*, 1988; Pocock, 1983; Meinert, 1986).

One argument against restrictive selection is that it may prevent generalisation of the results to the target population (Bailey, 1994; Cook and Campbell, 1979). However, lack of generalisation in studies with a very narrow population is not always a problem. It might well be possible to generalise results achieved in a small sub-group to a target population if a plausible explanation can be given (Davis, 1994). For instance, Lehman and Stonebridge (1978) have concluded that an intelligent knee unlocking system will not be viable during paraplegic walking because of the low walking speed. Though the patients in that particular study had low level lesions, the results may also be generalised to high level paraplegics, since their walking speed is, in general, lower.

Conclusions and recommendations

Most of the studies which were included in this survey have conducted a simple within subject comparison without randomisation of the order. These designs are considered internally invalid as they do not provide the possibility to control for aspecific treatment effects. Randomisation of the measurement order, which ultimately provides a cross-over design, is considered essential in order to improve the internal validity as the aspecific effects (period-effects) can be controlled.

All studies lack statistical power due to the small sample sizes and the heterogeneity of the study population. While interval estimation may improve interpretation of (statistically) negative trials, it appeared that there still was insufficient precision to conclude whether the differences were clinically relevant. Two recommendations can be put forward, viz. include more patients and apply relevant inclusion and exclusion criteria in order to obtain a more homogeneous study sample. Though restriction of the study population prevents generalisation, it is far more important to conduct a trial which is internally valid and provides precise estimates of the differences.

Different alternatives can be advocated if it appears impossible to include more patients, including multi-centre trials, statistical pooling of different small studies (i.e. meta-analysis) and other methodological approaches.

Though different methodological approaches should not be considered as a first choice, interrupted time series may offer some advantages. In an interrupted time series, which is essentially an extension of the AB/BA cross-over design, one obtains more statistical power because repeated measurements of the same subject are included in the analysis (Wagenaar, 1990). Moreover, because the analysis is performed on a single subject, heterogeneity of the study population is no longer affecting the statistical power.

REFERENCES

ALTMAN DG (1980). Statistics and ethics in medical research. III How large a sample? *Br Med J* **15 November**, 1336-1338.

ALTMAN DG, BLAND JM (1995). Absence of evidence is not evidence of absence. *Br Med J* **19 August**, 485.

BAARDMAN G, IJZERMAN MJ, LEERDAM NGA VAN, HERMENS HJ, VELTINK PH, BOOM HBK AND ZILVOLD G (1997). The SEPRIX: a separable reciprocator with intelligent knee stabilisation for people with paraplegia. In: Design and evaluation of a hybrid orthosis for people with paraplegia. PhD Thesis. – Enschede: University of Twente.

BAILEY KR (1994). Generalizing the results of randomised clinical trials. *Controll Clin Trials* **15**, 15-23.

BARNARD GA (1990). Must clinical trials be large? the interpretation of p-values and the combination of test results. *Stat Med* **9**, 601-614

BEILLOT J, CARRÈ J, LE CLAIRE G, THOUMIE P, PERRUOIN-VERBE B, CORMERAIS A, COURTILLON A, TANGUY E, NADEAU G, ROCHCONGAR P, DASSONVILLE J (1996). Energy consumption of paraplegic locomotion using reciprocating gait orthosis. *Eur J Appl Physiol* **73**, 376-381.

BORENSTEIN M (1994). The case for confidence intervals in controlled clinical trials. *Controll Clin Trials* **15**, 411-428.

BULPITT CJ (1987). Confidence intervals. *Lancet* **28 February**, 494-497.

CARPENTER LM (1993). Is the study worth doing? *Lancet* **24 July**, 221-223.

COOK TD, CAMPBELL DT (1979). Quasi-experimentation: design and analysis issues for field settings. – Boston: Houghton Mifflin.

DALLAL GE (1990). PC-size: consultant - a program for sample size determinations. *Am Stat* **44**, 243.

DAVIS CE (1994). Generalizing from clinical trials. *Controll Clin Trials* **15**, 11-14.

DUPONT WD, PLUMMER WD (1990). Power and sample size calculations. A review and computer program. *Controll Clin Trials* **11**, 116-128.

EDWARDS J, BATAWEEL AOP (1996). Hybrid system for upright mobility with unlockable orthotic knee for knee bending during swing phase. In: Neuroprosthetics from basic research to clinical applications. – Berlin: Springer Verlag. p523-530.

FEINSTEIN AR (1985). Clinical epidemiology. The architecture of clinical research. – Philadelphia: W.B. Saunders.

FERRARIN M, RABUFFETTI M (1996). On the improvement provided by hip transversal rotation on paraplegic gait walking in reciprocating gait orthosis. In: biomedical and health research program (Bio Med) of the European Union. In: Neuroprosthetics from basic research to clinical applications./edited by A Pedotti, M Ferrarin, J Quitren, R Reiner. – Berlin: Springer Verlag. p493-502.

FREEMAN PR (1993). The role of p-values in analysing trial results. *Stat Med* **12**, 1443-1452.

GALLIEN P, BRISSOT R, EYSSETTE M, TELL L, BARAT M, WIART L, PETIT H (1995). Restoration of gait by functional electrical stimulation for spinal cord injured patients. *Paraplegia* **33**, 660-664.

GARDNER MJ, ALTMAN DG (1986). Confidence intervals rather than p-values: estimation rather than hypothesis testing. *Bri Med J* **15 March**, 746-750.

HIROKAWA S, GRIMM M, LE T, SOLOMONOW M, BARATTA RV, SHOJI H D'AMBROSIA RD (1990). Energy consumption in paraplegic ambulation using the reciprocating gait orthosis and electrical stimulation of the thigh muscles. *Arch Phys Med Rehabil* **71**, 687 - 694.

ISAKOV E, DOUGLAS R, BERNS P (1992). Ambulation using the reciprocating gait orthosis and functional electrical stimulation. *Paraplegia* **30**, 239 - 250.

JEFFERSON RJ, WHITTLE MW (1990). Performance of three walking orthoses for the paralysed: a case study using gait analysis. *Prosthet Orthot Int* **14**, 103-110.

KLEINBAUM DG, KUPPER LL, MULLER KE (1988). Applied regression analysis and other multivariable methods./2nd edition, – Belmont, California: Duxbury Press.

LACHIN JM (1981). Introduction to sample size determination and power analysis for clinical trials. *Controll Clin Trials* **2**, 93-113.

LEHMAN J, STONEBRIDGE J (1978). Knee lock device for knee ankle foot orthoses for spinal cord injured patients: an evaluation. *Arch Phys Med Rehabil* **59**, 207-211.

LOTTA S, FIOCCHI A, GIOVANNINI R, SILVESTRIN R, TESIO L, RASCHI A, MACCHIA L, CHIAPATTI V, ZAMBELLI M, TOSI C, BARATTA S, FRANCESCHINI M (1994). Restoration of gait with orthoses in thoracic paraplegia: a multicentric investigation. *Paraplegia* **32**, 608-615.

MCCLELLAND M, ANDREWS BJ, PATRICK JH, FREEMAN PA, EL MASRI WS (1987). Augmentation of the Oswestry parawalker orthosis by means of surface electrical stimulation: gait analysis of three patients. *Paraplegia* **25**, 32-38.

MEINERT CL (1986). Clinical trials: design conduct and analysis. – New York: Oxford University Press

MUSZKAT R, YAZBEK P, ARANGO CT, MOREIRA MC, TROMBETTA IC, BATTISTELLA LR (1994). Assessment of functional capacity during gait using a reciprocal propulsion orthosis (ARGO) - a comparative study with a conventional mechanical orthosis. *Sao Paulo Med J* **112**, 495-499.

NENE AV, PATRICK JH (1989). Energy cost of paraplegic locomotion with the ORLAU parawalker. *Paraplegia* **27**, 5-18.

NENE AV, PATRICK JH (1990). Energy cost of paraplegic locomotion using the parawalker-electrical stimulation "hybrid orthosis". *Arch Phys Med Rehabil* **71**, 116-120.

NENE AV, JENNINGS SJ (1992). Physiological cost index of paraplegic locomotion using the Orlau parawalker. *Paraplegia* **30**, 246-252.

OTTENBACHER KJ (1995). Why rehabilitation research does not work (as well as we think it should). *Arch Phys Med Rehabil* **76**, 123-129.

OTTENBACHER KJ, BARRETT KA (1990). Statistical conclusion validity of rehabilitation research. A quantitative analysis. *Am J Phys Med Rehabil* **69**, 102-107.

PETROFSKY JS, SMITH JB (1991). Physiologic costs of computer-controlled walking in persons with paraplegia using a reciprocating gait orthosis. *Arch Phys Med Rehabil* **72**, 890-896.

PHILLIPS CA (1989). Electrical muscle stimulation in combination with a reciprocating gait orthosis for ambulation by paraplegics. *J Biomed Eng* **11**, 338-344.

POCOCK SJ (1983). Clinical trials, a practical approach. – Chichester: John Wiley.

POLLOCK C, FREEMANTLE N, SHELDON T, SONG F AND MASON JM (1993). Methodological difficulties in rehabilitation research. *Clin Rehabil* **7**, 63-72.

REILLY RP, FINDLEY TW (1989). Research in physical medicine and rehabilitation. IV, Some practical designs in applied research. *Am J Phys Med Rehabil* **68**, 196-201.

ROTHMAN KJ (1986). Modern epidemiology. – Boston: Little Brown and Company.

SACKS HS, BERRIER J, REITMAN D, ANCONA-BERK VA, CHALMERS TC (1987). Meta-analyses of randomized controlled trials. *New Engl J Med* **316**, 450-455.

SAITOH E, SUZUKI T, SONODA S, FUJITANI J, TOMITA Y, CHINO N (1996). Clinical experience with a new hip-knee-ankle-foot orthotic system using a medial single hip joint for paraplegic standing and walking. *Am J Phys Med Rehabil* **75**, 198-203.

SALSBURG D (1990). Hypothesis versus significance testing for controlled clinical trials: a dialogue. *Controll Clin Trials* **9**, 201-211.

SENN S (1993). Cross-over trials in clinical research. – Chicester: John Wiley.

SMITH WDF, CLARK PF, MACARTHUR D, ALLATT RD, HAYES KC, CUNNINGHAM DA (1997). Oxygen cost using a reciprocating gait orthosis in a paraplegic (T9) patient with a bilateral below-knee amputation: case report. *Spinal Cord* **35**, 121-123.

SMITH AH, BATES MN (1992): Confidence limit analyses should replace power calculations in the interpretation of epidemiologic studies. *Epidemiology* **3**, 449-452.

STALLARD J, MAJOR RE (1995). The influence of orthosis stiffness on paraplegic ambulation and its implications for functional electrical stimulation walking systems. *Prosthet Orthot Int* **19**, 108-114.

SYKES L, CAMPBELL IG, POWELL ES, ROSS ERS, EDWARDS J (1996). Energy expenditure of walking for adult patients with spinal cord lesions using the reciprocating gait orthosis and functional electrical stimulation. *Spinal Cord* **34**, 659-665.

THOUMIE P, PERROUIN-VERBE P, LE CLAIRE G, BEDOISEAU M, BUSNEL M, CORMERAIS A, COURTILLON A, MATHE JF, MOUTET F, NADEAU G, TANGUY E, BEILLOT J, DASSONVILLE J, BUSSEL B (1995). Restoration of functional gait in paraplegic patients with the RGO-II hybrid orthosis. A multicentric controlled study. I Clinical evaluation. *Paraplegia* **33**, 647-653.

WAGENAAR RC (1990): Controlled single case experimental design: the rehabilition of stroke patients. In: Functional recovery after stroke. – Amsterdam: Free University, University Press.

WHITTLE MW, COCHRANE GM (1989). A comparative evaluation of the hip guidance orthosis (HGO) and the reciprocating gait orthosis (RGO). *Health Equipment Information* **192**.

WINCHESTER PK, CAROLLO JJ, PAREKH RN, LUTZ LM, ASTON JR JW (1993). A comparison of paraplegic gait using two types of reciprocating gait orthoses. *Prosthet Orthot Int* **17**, 101-106.

YANG L, GRANAT MH, PAUL JP, CONDIE DN, ROWLEY I (1996). Further development of hybrid functional electrical stimulation orthoses. *Spinal Cord* **34**, 611-614.

YUSUF S, COLLINS R, PETO R (1994). Why do we need some large, simple randomised trials? *Stat Med* **3**, 409-420.

IJZERMAN MJ, BAARDMAN G, HERMENS HJ, VELTINK PH, BOOM HBK, ZILVOLD G (1997[a]). The influence of the reciprocal cable in the advanced reciprocating gait orthosis on paraplegic gait performance. *Prosthet Orthot Int* **21**, 52-61.

IJZERMAN MJ, BAARDMAN G., HOLWEG GGJ, HERMENS HJ, VELTINK PH, BOOM HBK AND ZILVOLD G (1997[b]). The influence of frontal alignment in the advanced reciprocating gait orthosis on energy cost and crutch force requirements during paraplegic gait. *Basic Appl Myol* **7**, 123-130.