

Project Number: DE3101 **Project Acronym: EATS**
Project title: Efficiency of Assistive Technology and Services
Deliverable Type: PU

Deliverable Number: D2.3
Contractual date of delivery: June 30, 1997
Actual date of delivery: August 29, 1997
Title of deliverable: Report on prototype instruments
Work package contributing to the deliverable: WP02
Nature of deliverable: RE
Authors: [Håkan Brodin](#) [Jan Persson](#) (CMT, Sweden)
[Roelof Wessels](#) (IRV, The Netherlands)

Abstract:

Areas for policy making concerning assistive technology and services (AT&S) and the focus of EATS are discussed related to the merits of outcome measures such as psychometric profiles, goal attainment scales, the problem elicitation technique, and utilities.

It is proposed that the methods to be chosen in EATS are goal attainment scaling or the problem elicitation technique, and a utility measure. The utility measure should primarily be based on the EuroQol instrument revised as proposed in the previous TIDE project CERTAIN. This means that aspects on disability and handicap should be included and health aspects excluded.

Keyword list:

outcome measures, efficiency, cost-effectiveness, assistive technology

EXECUTIVE SUMMARY

Areas for policy making concerning assistive technology and services (AT&S) and the focus of EATS are discussed related to the merits of outcome measures such as psychometric profiles, goal attainment scales, the problem elicitation technique, and utilities.

It is proposed that the methods to be chosen in EATS are goal attainment scaling or the problem elicitation technique, and a utility measure. The utility measure should primarily be based on the EuroQol instrument revised as proposed in the previous TIDE project CERTAIN. This means that aspects on disability and handicap should be included and health aspects excluded.

CONTENTS

1.	INTRODUCTION	4
	1.1 Background	
	1.2 Economic aspects on measurement practice	
2.	PSYCHOMETRIC PROFILES, GOAL ATTAINMENT SCALES (GAS) AND UTILITIES	7
	2.1 Categories of instruments	
	2.2 Original GAS instruments	
	2.3 Distinction between GAS and utilities	
3.	GOAL ATTAINMENT SCALING FOR AT&S	13
	3.1 GAS experiences in rehabilitation	
	3.2 Guidelines for GAS use in assistive technology	
	3.3 Problem elicitation technique (PET) or GAS?	
4.	UTILITY MEASUREMENT FOR AT&S	19
	4.1 Time trade-off (TTO) in practice	
	4.2 Guidelines for TTO use in AT&S	
5.	CONCLUSIONS	26
	REFERENCES	27

1. INTRODUCTION

1.1 Background

This report describes the methodological basis for outcome instruments intended to be developed and used in the EATS study. From this basis, instruments and manuals for their utilisation will be developed.

The methodology to be considered follows three lines:

- goal attainment scaling (GAS),
- problem elicitation technique (PET), and
- utility measures.

All are genuinely based on end-users' opinions and preferences. The methodology has been considered in national workshops held during April and May 1997 in the countries of the EATS partners (Sweden, The Netherlands, Norway and Italy) with participation of end-users, user organisations, policy makers and administrators, professionals, experts on outcome measurements and industry. The proposed methodology was supported by these workshops, although the GAS approach means a new road when it is applied to the field of assistive technology and services (AT&S).

The background to the EATS development of outcome measurements in AT&S was laid in the CERTAIN study in the TIDE program (1994 - 1996). In this study it was pointed out that existing methods for outcome measurements, i.e. general health profiles as well as utility measures, had deficiencies and that refinements have to be made for appropriate assessments of AT&S (Andrich and Ferrario 1996, Lorentsen and Hem 1996, Oortwijn and van Beekum 1996a, 1996b, Persson and Brodin 1996, Persson et al. 1997).

1.2 Economic aspects on measurement practice

When assessments of user preferences or goal fulfilment are done for assistive technology and services, for economic or policy purposes, it is important that the measurement is in accordance with accepted general economic framework. Such basic theoretical prerequisites have been reviewed in the study HEART (1992 - 1994) of the TIDE program (Persson and Brodin 1993, Alastuey et al. 1994a, 1994b) and the CERTAIN study. It was found that the general methods for socio-economic evaluation also were applicable in the AT field, but that the health oriented perspective must be changed towards an ability oriented perspective. However, there are also some prerequisites of the measurement practice which was not covered in the previous studies. This section presents the results of a literature review

undertaken to identify the necessary economic methodological framework, which should be followed when using goal attainment (GAS) and time-trade-off (TTO) techniques in practice. The section, to be of assistance in the following field testing of the instruments, is divided in two, one for each outcome measure, viz. GAS and utility measures.

One of the conclusions from the CERTAIN study was that it might be necessary to use different ways of measuring outcomes depending on the purpose of the assessment (Persson and Brodin 1996). EATS recognises five major areas, see Table 1, which are addressed by decision makers, where outcomes analysis provides important information

Table 1. Major areas for decision making.

-
1. Technological development
 2. Priorities between AT programmes
 3. Quality assessment programmes for service providers
 4. Professional guidelines in clinics and service units
 5. Management of individual services (not covered by EATS)
-

Different areas might need different socio-economic techniques, such as cost-effectiveness (CEA) and cost-utility analysis (CUA) (Persson and Brodin 1996). Levels 3 and 4 have been judged to be most urgent to investigate, and two approaches were identified, the *goal attainment approach* and the *utility approach*, reflecting the CEA and the CUA perspectives respectively.

All levels would benefit from using the utility technique. However, levels 3 and 4 may also need a more specific outcome instrument to reflect the goal attainment of the services in a narrower sense, used in a cost-effectiveness analysis. Such instruments – goal attainment scales (GAS) – have been used for *individual* treatment programmes in mental hospitals and in the rehabilitation areas, but they have not been used for cost-effectiveness studies of AT on *policy* levels.

Before applying a health status instrument, the following issues should be addressed:

1. The purpose for which an instrument is used must be clearly stated.
2. The concepts to be measured should be clear for every instrument.
3. The instrument to be used must have certain attributes which will determine its usefulness.

There are three essential attributes that need to be addressed:

- reliability,
- validity or accuracy,

- responsiveness.

It is important to specify the viewpoint for which an instrument is used. Next to the determination of the viewpoint of interest the concepts to be used need to become obvious (determine critical outcomes). Next to this the following issues need to become clear:

- relevance with regard to critical factors,
- relevance of the questions used in the instrument,
- practicability and level of dissemination of the instrument,
- validity and reliability of the instrument.

2. PSYCHOMETRIC PROFILES, GOAL ATTAINMENT SCALES (GAS) AND UTILITIES

2.1 Categories of instruments

AT&S provides benefits for the clients in terms of reduced impairments, augmented abilities and reduced handicap. The goal of the prescription of an assistive device may address one or several of these levels. How well the goal is attained may be measured through psychometric profiles, or some type of goal attainment scale (GAS). This may serve as the effectiveness measure in a CEA. The individual's preferences of various situations is the integration and valuation of domains of the effectiveness concept by the client. The result may be used in a CUA. Thus, the first phase of a CUA may be an identification of such domains and assigning values to them (see Table 2).

Table 2. Types of instruments for rehabilitation outcomes (elaborated from Persson and Brodin 1996).

Generic instruments	Health Profiles	Examples: Nottingham Health Profile (NHP) Sickness Impact Profile (SIP) Short Form 36 (SF-36) GAS
	Utility measures	Classification into predetermined utility groups (decomposed approach) Rosser-Kind Index IHQL-The Index of Health-related Quality of Life 15D-The 15D-measure EuroQol-See also EuroQol Group (1990) QWB-The Quality of Well-being Scale MMHCS-The McMaster Health Classification System QUID -New utility instrument proposed by the EATS project
		Clients self-rating (holistic approach) Rating scale Standard gamble Time trade-off Willingness to pay
Disease-specific instruments		

The objectives of EATS are primarily to find outcome measures which may support major decision areas as to Table 1, items 2, 3 and 4, i.e., priority setting, quality assessment and professional guidelines. Results should be achieved on population level. Global measures adequate for comparative analyses are required. Certainly, psychometric profiles most probably

will yield non-comparable measures. GAS or similar measures (see PET below) and utilities provide universal measures, the usability of which will be investigated in EATS.

An issue is whether professionals and clients may have conflicting opinions. The client may have a limited understanding of, e.g., medical consequences like joint and muscle loads, of various prescriptions. Such issues puts the study design in focus. It is assumed that the services is provided according to scientific state-of-the-art, or that optimal praxis is searched for. An ex ante - ex post study may be used to provide clients' ultimate ratings/valuations.

In medical treatment it might be of great value to monitor the treatment program quality as well as individual case histories. Goal attainment scales (GAS) are well established within the nursing area to examine "numbers of single treatment cases". Some attempts are also made to evaluate the societal value, that is to assess the value not only for the single patients, but also for family and society. The designers of the GAS instrument emphasises that it has not been intended for policy purposes, but that it would be possible and valuable to do so (Kiresuk, Stelmachers and Schultz 1982, Kiresuk, Smith and Cardillo 1994). The EATS project intends to adapt the GAS instrument for use in the CERTAIN tool, for cost-effectiveness purposes on policy level. The method is well described in the textbook by Kiresuk et al. (1994).

2.2 Original GAS instruments

GAS has been used for evaluation of individual treatment programs, originally in mental care. It was developed by Kiresuk and Sherman (1968) and was adapted and used in several areas where outcome measurement is especially complex, for example in psychotherapy, sociotherapy and physiotherapy.

If the GAS technique is to be adopted for outcome measures in an economic framework it has to be certified that its properties conform with already existing instruments, so that results from different cost-effectiveness studies are commensurable. A critical issue in this process is how the *goal* concept is to be understood.

The HEART and CERTAIN studies recognised the concepts of productivity and efficiency as two different ways of expressing the outcomes related to costs (Persson and Brodin 1996). The difference was demonstrated in a simple statement: "High productivity is to do something right, but high efficiency is to do the right thing". That is, in many cases the productivity may be high but the efficiency is poor, because the process is not in the direction of the goal. Thus, the focus of the EATS study is on effectiveness and/or efficiency and not on productivity and the outcome measurement should be addressed accordingly.

Goals in medicine may be analysed in several ways (Liss 1990). In economic theory *goals* (related to efficiency) are not the same as *outcomes* (related to productivity). The term *Goal fulfilment* or *goal attainment* is not only a numerical expression of the desired degree of attained outcomes, but it describes outcomes in respect to the available resources used. It is without meaning to mention goals without also expressing the means to reach these goals. The economic interpretation of a *goal* is then essential to the use of effectiveness and efficiency measures. When *cost-effectiveness* is the focus of the study (for instance in a CEA), the outcomes are measured in physical units, and related to the costs, e.g. the resource use. When *cost-utility* is considered (for instance in a CUA), instead the value of the outcomes, e.g. measured in QALYs or HYE, are related to costs.

It is somewhat unclear how to deal with unintended or unwanted outcomes. In medicine, and especially in pharmaceuticals, it is rather a rule than an exception that important medical side effects appear. These unwanted effects must also be considered when the drugs are evaluated. Should they appear as a result – an outcome, or should they be considered as a sacrifice on the resource side, which we have to accept to receive what we actually want. The same discussion applies to assistive technology. Should unintended effects, for instance negative feelings from using stigmatising devices, or wounds from tight-fitted prostheses, be considered as the type of sacrifices above on the cost side, or as a negative outcome on the outcome side? No clear answer could be given, but the CERTAIN project established a clarifying principle that all changes in resource use, positive or negative, should be collected to the cost side, thus following the principle that a goal could never be to cut or increase resources. Resources are always means to attain goals. This implies, in turn, that goals must be expressed in other terms, for instance in quality of life measures (which are clearly no resources).

2.3 Distinction between GAS and utilities

All socio-economic analysis may be performed through the process of 1. *Identification*, 2. *Quantification* and 3. *Valuation* of costs and outcomes. Value estimations are done by assigning preference based weights (i.e. values) to the quantities of the second phase. The GAS instrument differs from utilities in that it only goes through the first two steps. However, the possibility of weighting goals means a kind of valuation. This valuation does not mean that its result is equal to or reflects the utility. (It seems possible to perform the weighting in a way so as to achieve utilities, e.g. by means of time trade-off or standard gamble techniques. This idea needs further elaboration.)

This fundamental property of the GAS instrument implies an important limitation in the applications. Many instruments recognise *multiple* outcomes and so does GAS. The designers advocate at least three goals (Kiresuk et al. 1994, p.8). The GAS instrument then could be classified into the generic, health profile group of Table 2 (Persson and Brodin 1996).

It was recognised that in some cases health profiles, measuring a number of aspects of quality of life, may be weighted together into a single score (e.g., Sickness Impact Profile). However, to be able to do so an extensive validation process must be applied and there is also a trade-off between sensitivity and generality which might make them inadequate for evaluation of certain program and technology developments (Hopkins 1992, Kaplan 1993). Often the designers of outcome instruments strive to measure aspects of quality of life along the three physical, mental and social dimensions, and to get those dimensions as independent as possible of each other (orthogonality) to avoid multicollinearity (Kiresuk et al. 1994, p.191). When it comes to multiattribute GAS goal setting, there is no requirement that the different goals should be as different as possible. That implies that the evaluator can not be sure of what algorithms could be used for aggregation to the single score. The designers of the original GAS instrument state that the aggregation procedure could be chosen from a variety of techniques (Kiresuk et al. 1994, p.177) and they consequently also strongly advice users not to weigh the different dimensions (Kiresuk et al. 1994, p.192). To conclude: the applicability for *multiattribute* GAS assessments is limited without going through the valuation phase, since the GAS is unable to weigh different goals except in an arbitrary sense.

The GAS application is still of great value as a cost-effectiveness instrument since some properties are of particular interest for economic evaluations. The professional's judgement is used before the treatment to form a scale against which the progress is assessed, to achieve an objective measurement. But the scale also serves as a guide against which treatment activities are focused. The individual patient participates both in the definition of the goals and in the assessment of the results. That means that the user influence should be high, given that the goal

setting procedure is designed to promote equal decision power from user and provider. A small warning may be issued, however. It seems like the idea of GAS is beguilingly simple. The designers of the instrument stress at several opportunities that training in using GAS is necessary to make it functional (Kiresuk et al. 1994, Chapter 5, Choate et al. 1981).

However, some properties of the instrument need to be scrutinised. The original instrument is constructed from an anticipated outcome as the expected result. Results above this level will give a positive scale value, results below will give a negative value. This is somewhat different to the economic way of defining goals. A positive characteristic is that the ways of defining the outcomes is clearly established in the original GAS instrument. The scale, with an outcome range from -2 to +2 is clearly ordinal in its nature. The value 0 is given an expected result, the value +2 is given to a result much better than expected. This defines a range of universal understanding but on the other hand the use of the scale values is essential to how the evaluation expression will appear. The proposed expression is sensitive to the number of goals. The results with only one goal ranges between 30 for "much less than expected" and 70 for "much better than expected". If five different goals are set with equal weights results range from 19.8 to 80.2. Different weights for different goals will also alter the results.

Some differences are found when GAS is to be included in common cost-effectiveness analysis. In CEA it is customary that participating professionals do not actively take part in the assessment to reduce the influence of evaluation bias. This has both positive and negative consequences. For research projects the required objectivity and general validity is not so difficult to defend if the assessment is done outside the project. But in routine treatment situations the service provider actively participates in the progress and his or her knowledge of the situation is also often essential to the treatment result.

A second difference is that scales in economic evaluations typically are chosen from external and universally known sets of criteria. An example is the measurement of diastolic blood pressure in the clinical trial of a new hypertension drug, where one millimetre is the same regardless where the measurement is done. The original GAS scale, on the other hand, needs not to be externally transparent. Furthermore, economic analysis has the advantage of not only setting defined end-goals but allows asymptotically non-attainable states by using interval or ratio scales. An example of the latter is, once again, evaluation of hypertension treatment where the ultimate but non-realistic (asymptotic) goal is to reduce blood pressure to normal. The proper question rather is to find the optimal degree of goal fulfilment. This is only possible if both cost and outcome scales use equidistant scale steps.

The conclusion is that the original GAS:

1. provides a global measurement of goal fulfilment without using preference values,
2. allows the user to express his or her views, also without utility assessment,
3. can not be used for external evaluation unless universal measures is used,
4. the scale is not globally commensurable unless interval or ratio scales are used.

3. GOAL ATTAINMENT SCALING FOR AT&S

According to the conclusions above the original GAS scale is modified to a scale measuring goal fulfilment *in percentage of the pre-set goal*. In this way the conclusion 3 and 4 above are fulfilled. We may now use the common apparatus for a) assessing total scores from a set of sub-goals and b) comparing the results from the assessment with other studies within the service unit or program. It is assumed that the measurement precision using percentage scores is comparable to the precision of the original GAS scale (-2 to +2). However, any potential serious drawbacks by changing the scale has yet to be evaluated through empirical field tests.

3.1 GAS experiences in rehabilitation

Three examples will be reported which may contain experiences which are valuable in the development of the EATS GAS. The examples are all from the field of occupational therapy.

1. The Canadian Association of Occupational Therapists have, in collaboration with Health and Welfare Canada, developed an outcome measure, The Canadian Occupational Performance Measure (COPM). COPM is an outcome measure which is client-centred and encompasses self-care, productivity and leisure (Law et al. 1990). COPM yields two scores: performance and satisfaction with performance.

COPM is administered in five steps: 1) problem definition, 2) problem weighting, 3) scoring, 4) re-assessment, and 5) follow-up.

The methodology yields essential experiences as to how to identify a problem, how to rate the importance of activities, how to rate the ability to perform these activities (a 1-10 scale), how to aggregate the scores of the various activities, and experiences from measurement of change (related to intervention). Finally, reassessment is used for planning purposes of continued treatment or prescription.

The experiences cover feasibility studies, also dealing with problems as timing of the instrument administration, the use with care givers, and differences in opinions between clients, therapists and family members. It was concluded that COPM has a potential for programme evaluation and quality assurance in populations with similar problems. A caution is raised with regard to comparisons between groups with little similarity in the problems identified.

2. Penny Spreadbury (1997) reported on a binary individualised outcome measure used at Nottingham City Hospital NHS Trust. She notes that instruments as Barthel Index and SF-36 do not take clients' opinions or goals for treatment into account. She uses a simple binary

outcome measure where each problem or goal is scored with "yes" or "no" (1 or 0) depending on whether the expected outcome was achieved or not.

A three-stage process is used: 1) individualised problems or goals are identified and negotiated with the client, 2) a programme is planned and implemented, and 3) after an agreed time actual achieved outcomes are compared with the expected outcomes to measure whether the programme was effective.

The method has been successful in defining outcomes including daily living skills, mobility, use of equipment, social behaviour, knowledge, feeling and attitudes. Feelings and attitudes are most important in a holistic approach to rehabilitation, the author claims. It is concluded that the binary outcome measure is useful for audit and service review and outcomes monitoring purposes. "It is client centred, emphasises a holistic therapeutic approach and recognises the autonomy of the individual client."

3. At the departments of rehabilitation, neurology and geriatrics of the Linköping University Hospital a program for quality management and outcomes monitoring of occupational therapy has been on-going since one year. The experience is that GAS per se is difficult to use. Instead, the process was broken down into nine issues: 1) is the problem defined, 2) is a goal identified and determined, 3) is the goal measurable, 4) is the goal on a functional level, 5) is the goal on an ability level, 6) have measures taken been documented, 7) is the goal assessed, 8) is the goal attained, 9) is there a relation between problem, goal and outcome.

It is noted that more work on goal assessment is needed. A short memo in Swedish was given by Samuelsson and Wressle (1996).

3.2 Guidelines for GAS use in assistive technology

We adapt the guidelines from Kiresuk et al. (1994 p.7) to the AT field and percentage scales:

1. Identify the issues that will be the focus of assistive technology.
2. Translate the selected problems into at least three goals.
3. Choose a brief title for each goal.
4. Select an indicator for each goal.
5. Specify the maximum level of outcome for the goal.
6. Repeat these scaling steps for each of the three or more goals.

3.3 Problem Elicitation Technique (PET) or GAS?

Introduction

This section will give some comments on the choice of GAS and a possible alternative. After gaining a bit more in-depth knowledge about Goal Attainment Scaling (GAS), we were getting the impression that GAS might not be the most appropriate method for the level 3 and 4 (quality assessment and professional guidelines respectively) objectives of EATS after all. The conditions under which GAS leads to reliable measurements seem to conflict with the context in which the effectiveness instrument is to be applied during (and after) the EATS project. These comments will start with a short description of GAS. Next the objections against GAS with respect to the particular application within EATS will be pointed out. Finally a short description will be given of what might possibly be a suitable alternative: the so-called Problem Elicitation Technique (PET).

Goal Attainment Scaling (GAS)

Goal Attainment Scaling (GAS) was first described as a general method to evaluate the outcome of mental health treatment. It consists of a strict procedure to structure and evaluate the goals of individual clients in the service delivery process. The method can be applied in every situation where one needs to analyse problems, specify goals and evaluate the degree of goal attainment afterwards. GAS is an instrument that allows you to make these actions explicit and to execute them systematically and objectively. GAS is being used in fields such as mental health care, education, public service delivery, management and research. By using GAS one does not gain insight in the causes of changes. GAS can only point out the degree of change with respect to expectations.

The GAS-method consists of the following steps. The problem of a client is dissected into specific sub-problems. For every sub-problem one or more goals are set and for each goal an indicator is selected. All indicators are to consist of descriptions of the next five levels of goal attainment:

- Much less than expected level of outcome.
- Somewhat less than expected level of outcome.
- Expected level of outcome.
- Somewhat more than expected level of outcome.
- Much more than expected level of outcome.

Constructing and evaluating the indicators is done according to a strictly defined procedure and by using a standardised evaluation form. Whenever possible, indicators have to consist of quantifiable variables. Different indicators can be ascribed a different weight. An evaluation period is agreed upon. Before the evaluation period, the level of goal attainment at which the client finds him- or herself for every sub-problem is scored on the evaluation form. After the evaluation period this is repeated; the difference indicates the level of goal attainment realised during the evaluation period. It is also possible to calculate an index value for both the situation at baseline and at follow-up, allowing for the difference to be quantified.

Objections against GAS

When applying GAS, in fact one measures the difference between the expected level of outcome and the level of outcome realised in reality. When the expected level of outcome is predicted perfectly, every measurement results in a T-score of 50. A differing T-score means a differing level of actual outcome and/or a poor prediction of the level of outcome. In order to guarantee a reliable prediction of the level of outcome and a proper formulation of indicators, certain conditions have to be met:

- Goal setters and follow-up interviewers should attend a proper training programme.
- Formulated goals and indicators should be monitored by peer review. By judging peers and being judged by peers, professionals need to optimise their ability to predict outcome.
- A substantial breaking-in period is required; goal setters need to acquire experience. Only after a substantial period of using GAS in a routine manner, it can become a reliable instrument.

This means GAS should not be considered a ready-made instrument that can be applied ad random every now and then by an interviewer. It is an instrument that can result in reliable measurements only when it is applied in a routine manner, when it is incorporated in the

organisation and its procedures and when the conditions of proper training, a peer review committee and a substantial breaking-in period are met. However, in EATS we are looking for an instrument that can be easily and quickly applied whenever a question arises. All in all, a reliable application of GAS requires a much more substantial adaptation of the organisation (and its procedures) in which it is applied, than intended in the EATS objectives.

Problem Elicitation Technique (PET)

An alternative method can possibly be found in the so-called Problem Elicitation Technique (PET). PET is, like GAS, a method for measuring those aspects that are relevant to the individual subject. With PET the disability status, as perceived by the subject, is assessed at baseline. After an intervention, this disability-status is measured again, so the difference can be assessed. The same measurement is done at two different moments in time and the results are compared. With GAS on the other hand, two different things (a prediction and an actual level of outcome) are measured and compared. The main objection against GAS, the reliability of the predicted level of outcome, does not apply to PET. Administering a PET questionnaire is done in four steps:

1. Identifying problems

The subject is asked to consider daily routine problems that he has been experiencing during the last week, as a result of his disability. Then he is asked to identify those problems that are most important to him and that he would like to see improved. Once the subject has finished identifying problems spontaneously, the interviewer reads him a checklist with possible problem areas. Subjects are allowed to identify up to 15 problems.

2. Scoring the impact of each problem from the subject's point of view

For each problem the subject is asked to fill out a form showing 4 Likert scales. The subject is to score one of the first three Likert scales for each problem, to indicate, depending on the nature of the problem, the level of difficulty, the degree of severity or the frequency of occurrence of the problem. The scales run from 0 to 7, equal to "without any difficulty" to "unable to do" (or "none" to "severe" or "never" to "always"). For each problem the subject is to score the fourth Likert scale, indicating the relative importance of the problem (with 0 equal to "least important" and 7 equal to "most important").

3. Summarising into a PET score

For each problem, the difficulty (or severity or frequency) score is multiplied by the importance score. The results are summed up and divided by the number of problems.

Therefore, the maximum score is 49, the minimum score is 0, a higher PET score indicating a higher degree of perceived disability.

4. Assessment of the subject's health score

The subject is asked to consider his overall health over the last week and indicate this on a numerical rating scale from 1 (worst possible health) to 10 (perfect health).

At follow-up the subject is asked to complete the same forms again, scoring the same problems (and health score) on the same scales. The subject notices his own previous scores and uses those as a framework of reference. New problems occurring between baseline and follow-up are not dealt with.

Discussion

Referring to what has been stated above we suggest to consider these methodologies further before a final decision on using the PET or the GAS or both. Methodologically, PET and GAS are not very different. However, PET avoids the reliability problems connected to the prediction procedures in GAS.

4. UTILITY MEASUREMENT FOR AT&S

The EATS project develops a new utility instrument, QUID, for CUA studies. The calibrating procedure (to achieve the predetermined QoL scores) of QUID uses the TTO technique (for a survey of the TTO technique, see Kaplan et al. 1993).

This assessment technique was developed as a result of the CERTAIN study. Since the principal outcome of AT often is quality of life aspects, the CUA technique would be the most appropriate among the three standard evaluation techniques (CBA, CEA, CUA). Utility/preference assessment may use two different methods for assigning values to the health or disability states. One, the holistic approach, uses clients' self-ratings based on hypothetical health state scenarios. In the decomposed approach, the respondents are asked to specify their state (health or disability condition). This may be done by means of a series of questions. Based on this, the respondents are assigned to categories with pre-determined utility scores. These scores have been previously determined by means of the holistic method, using a reference population.

Both of these approaches are used in QUID, the prototype instrument developed from CERTAIN results. QUID consists of two parts following the EuroQol line of practice (EuroQol Group 1990, O'Leary et al. 1995, Stalmeier et al. 1996), one that measures the holistic QoL through the thermometer rating scale type for instance used in EuroQol, and a second part which measures decomposed aspects according to a technique similar to IHQL but adapted for handicap related QoL (Persson and Brodin 1996).

4.1 Time trade-off (TTO) in practice

A growing number of studies use the QALY (quality adjusted life years) concept as an estimate of quality of life. The QALY is derived from weighting derived prolongation of life by the utility. A major part is produced using trade-off techniques against years of life, the TTO-technique. It may be anticipated that TTO applications will further increase as the concept will be known to a larger professional and analyst society, also to groups close to but outside the original health care area.

As was pointed out earlier TTO may be used for basically two purposes:

- 1) direct estimation of a person's (A) own current QoL condition (the holistic approach), and
- 2) hypothetically assessing somebody's (Bs) 'proxy' (Tsevat, Cook et al. 1994, Krahn et al. 1994, Walker and Rosser 1993 p.17) values of another person's (As) quality of life (the decomposed approach).

A third purpose, a mix of the two previous, could be found when previously assessed hypothetical conditions by the second technique (made by individual B) are determined for use in pre-valued QoL instruments of the own (As) current condition. This technique has been used in all instruments in the predetermined utility group of Table 2, and will also be used in the EATS project. We will use the term 'calibration' of the instrument for this process. The user of the AT does not have to perform the cumbersome and sometimes confusing process of several direct TTO estimates, but will state his or her own situation in a simpler identification, for later estimation by the predetermined values. CERTAIN produced guidelines for how QoL instruments should be designed and used in AT. The design of the instrument consists of two phases, basically following the identification, quantification and valuation principles of CERTAIN. However, using the TTO technique the quantification phase will be absorbed, partly in the identification phase and partly in the valuation phase. In this case the design of specific, separating and decisive questions also includes how to specify the different quantities of each aspect and this has been done in the identification phase. The valuing process, on the other hand, is not sensitive to different scale types because the TTO technique makes scaling implicitly into a ratio scale.

The direct TTO estimation technique has been presented earlier but the proxy calibration technique needs to be discussed. The available literature is not clear about how this should be done and the technique is not as easily standardised as the use of TTO itself. All varieties exist from a simple guess out of the air (Finlay et al. 1994) to very explicit, validated value systems from several groups of respondents (Stalmeier et al. 1996, Krumins et al. 1988). The EATS project uses a panel technique in two steps based on the view that there are different values in individual and group levels. Following the method of Krabbe et al. (1996) a group of experts, professionals and users responds to the proposed instrument (QUID), firstly individually (Session 1), secondly after an open discussion with only one united answer for the whole group (Session 2). In this way the validity is determined within the group and also the discussion may reveal aspects of the different questionnaire items not being obvious to all participants. Thus the end user influence will be more elaborated.

A large discussion is going on about the weighting of the individual items. One of the reasons of detailed questionnaires with individual scores is that several aspects of life could be pinpointed which maybe not are obvious to the respondent. The drawback of this technique is that the individual scores must be aggregated to one single score. Several aggregation techniques are available. Most of them origin from index techniques – averages of different kinds, arithmetic, geometric harmonic etc. A simple technique is also to sum the different scores and use them as deviations from full quality of life:

$$\text{QoL} = 1 - s \quad \text{where } s \text{ is the score of the individual item.}$$

Some authors argue that simple summation of individual items into one score will miss interaction between the items unless the items are completely independent from each other. Since this will be practically impossible and the interaction pattern also will vary between different respondents the approach is certainly not without errors. On the other hand there might be practical reasons for using a theoretically less defensible technique if the errors could be estimated or at least assessed in magnitude. Most QoL instruments take an interest in the whole spectrum of life from poor to excellent QoL. The summation technique above will be more erroneous the more items the respondent is marking in the instrument and, consequently, the probability of errors will be larger the worse the respondent's QoL is. However, presented studies from the health care area show that QoL is not an equally distributed function among the patient population but most respondents (even seriously ill) will be found in the upper part of the QoL spectrum, Figure 1 [see e.g. EuroQoL Group 1990, O'Leary et al. 1995, Tsevat et al. 1994). Consequently, the errors will be of some but minor importance for changes over time of an individual respondent. The probability of bias will be larger, however, in comparisons between users.

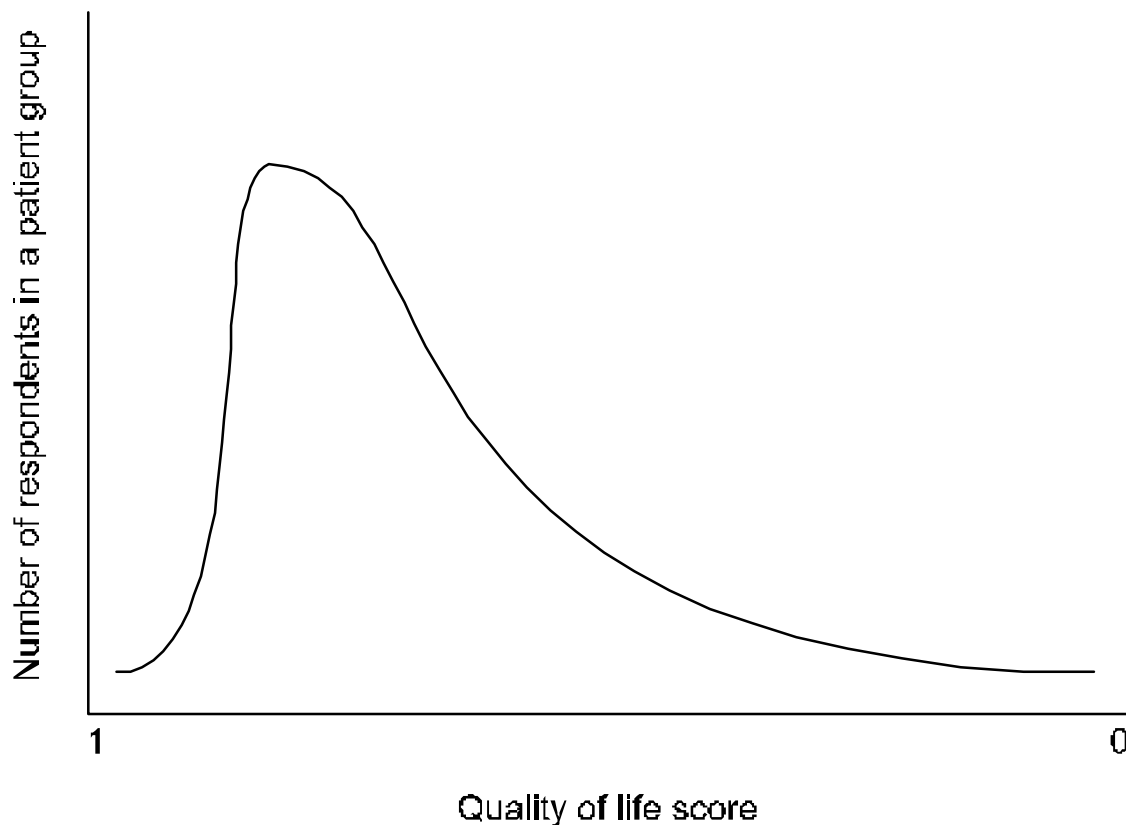


Figure 1. Distribution of QoL in a normal population.

To certify that the correct scores are attached to the items the panel consequently should make the assessments of session 2 in a matrix manner, so that different degrees of one score should be compared not only within the specific item but also to other items. The assessment should in principle follow the pattern of Figure 2 and a spreadsheet application is made for this purpose.

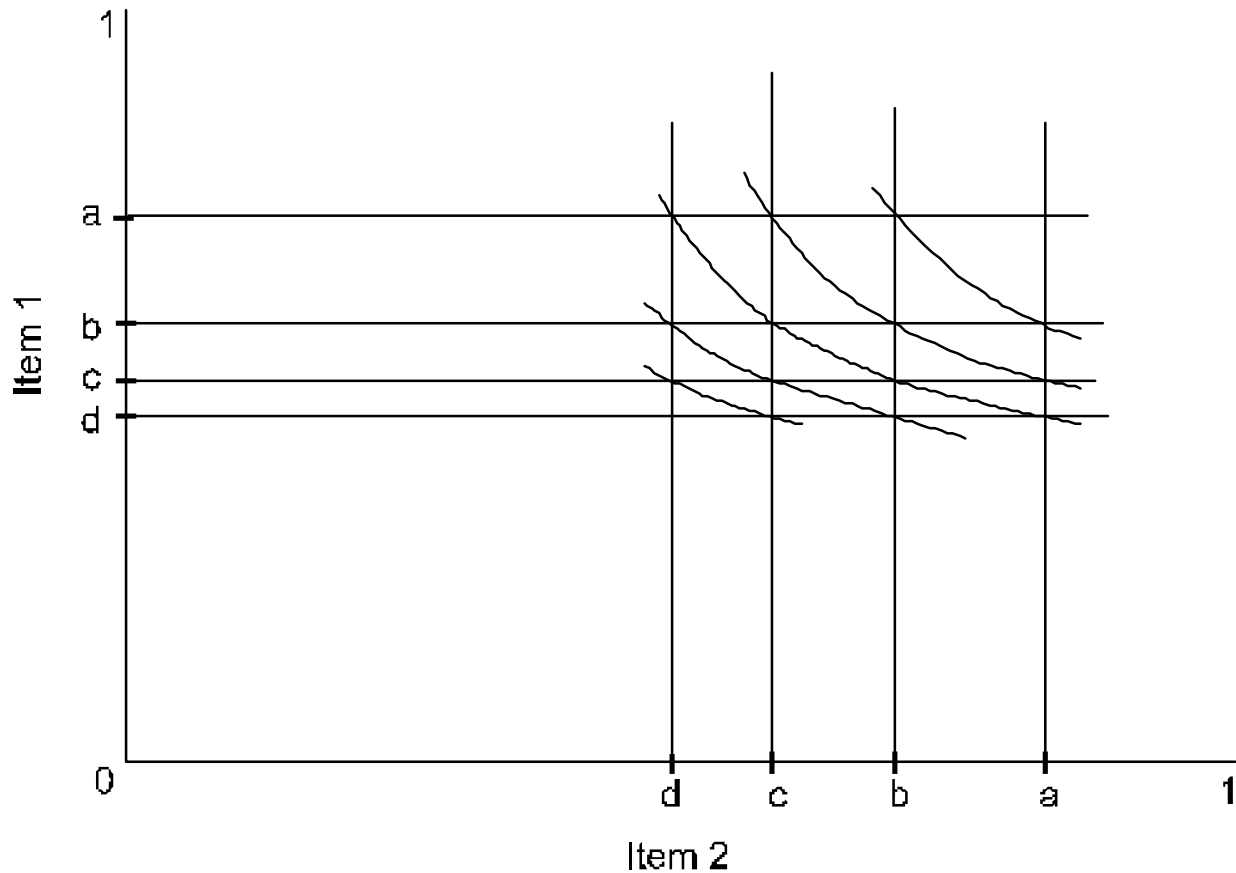


Figure 2. Indifference matrix of two QoL items. The curves are iso-QoL curves. Item 1 and item 2 are both factors which contribute to quality of life.

Each curve marks equal quality of life. The QoL decreases the closer the curves are to the origin. The slope of the indifference curves are arbitrarily chosen in the figure but the general topology is given. Since one indifference curve can not cross another, a graphic presentation of the scores of the respondents ensures that the assigned scores conform to the overall structure of the QALY matrix. The organiser of the assessment session 2 should follow the discussion and monitor the values in an indifference diagram, and present inconsistencies to the respondents to possibly alter values so that they will be in accordance with the general topic.

4.2 Guidelines for TTO use in AT&S

A form for TTO estimations in Session 1:

- a) An introductory text about the TTO technique and some definitions about for instance the difference between "moderate" and "mild" and so on.
- b) The description of the condition you want to estimate in QALY terms.
- c) The following tick test following several other TTO estimates (Tsevat et al. 1994).

Tick the appropriate question. Would it be better for you to live ·

Yes

Two healthy years or ten years with the described condition?

Two healthy years or seven years with the condition?

Two healthy years or five years with the condition?

Two healthy years or four years with the condition?

Two healthy years or three years with the condition?

Two healthy years or two and a half years with the condition?

Two healthy years or two years and one month with the condition?

Two healthy years or two years with the condition?

The tick test is repeated for each dimension of the instrument.

- d) Finally, information about personal characteristics is presented, including comorbidity, social and socio-economic data etc.

Session 2:

It may then be suitable in the second session to follow (and possibly adapt) the multi-attributive utility technique used for development of the IHQL instrument (Keeny and Raiffa 1976) to certify that the matrix will get the desired properties:

Stage 1: the levels of the descriptors are rated relative to the most severe level.

Stage 2: the descriptors are rated relative to each other on the appropriate scales.

Stage 3: the scales are rated relative to each other on the appropriate attributes.

Stage 4: the attributes are rated relative to each other on the appropriate dimensions.

Stage 5: the three dimensions are rated relative to each other.

Stage 6: the worst possible combination of all the descriptors (equivalent to extreme disability with extreme discomfort and extreme distress) are rated relative to being dead on a single scale.

Summary

To sum up, assessment of assistive technology may use the following structure to do a thorough outcome analysis for QUID:

1. Choose questionnaire items according to the procedure given in CERTAIN.
2. Create two assessment sessions consisting of experts, professionals and end users.
 - a) Session 1: Individual assessments of items and degrees of items without help in a TTO setting.
 - b) Session 2: Unified assessment guided by a chairperson, where the previous scores serve as departures for common agreements about a QoL indifference map, if necessary one map for each group of respondents.
3. Test the instrument validity against the thermometer instrument in a pilot of end users.

5. CONCLUSIONS

The final choice of methods for EATS is not yet taken. It is not possible to judge only from theoretical grounds how the methods work. Further consideration of pros and cons of GAS and PET respectively are needed using testing in practice. In developing instruments for GAS or PET and utility and manuals for their use the recommendations given above should be further elaborated, also taking the results from the EATS national workshops into consideration. While the development of utilities is based on experiences and recommendations from the study CERTAIN, the GAS and PET methods mean the implementation of a method with little experience available from the field of AT&S. The use of GAS in other areas shows, however, promising results.

The further development must consider a number of obstacles and problems:

- Who is an adequate guide for the client in the goal setting procedure?
- How to avoid goal setting guidance bias?
- How to score importance of goals and problems?
- How to avoid unrealistic goals (too high/too low)?
- How to overcome communication difficulties?
- What about comparisons between groups with dissimilar goals?
- What about comparisons between groups with dissimilar number of goals?
- Many people do not want or cannot take decisions themselves. How to interact in such situations with the clients?
- Are there ethical concerns about confidentiality, and are such concerns to be managed differently in different countries?

Although there are troublesome issues to solve, the GAS method is reported to be successful in many applications and means per se valuable attributes to the services. It promotes clients' involvement, it probably combines client and professional experience, it can take a holistic and multidisciplinary approach to the rehabilitation and may integrate clinical and psychosocial information. Finally, the client's perspective is completely decisive for the policy decisions and the client should become aware that his or her view is taken seriously.

REFERENCES

Alastuey J, Kerdraon M, Persson J, Brodin H, Petäkoski-Hult T Legal and macroeconomic factors impacting rehabilitation technology availability. D.2 Chapter 2: Assessment and refinement of existing socio-economic models. EC TIDE study Horizontal European Activities in Rehabilitation Technology. Commission of the European Communities, Directorate-General DG XIII, March, 1994a.

Alastuey J, Kerdraon M, Petäkoski-Hult T, Persson J, Brodin H Proposals for future actions on socio-economic models for studies of rehabilitation technology. Deliverable D.2 Chapter 3 in Line D: Legal and economic factors impacting rehabilitation technology availability. EC TIDE study Horizontal European Activities in Rehabilitation Technology. Commission of the European Communities, Directorate-General DG XIII, August, 1994b.

Andrich R, Ferrario M Cost outcome analysis for assistive technology: Case studies. EU TIDE study Cost-effective rehabilitation technology through appropriate indicators (CERTAIN). Commission of the European Communities, Directorate-General DG XIII, August 1996.

Bakker CH Patient-oriented outcome assessment in rheumatic diseases; Thesis Rijksuniversiteit Limburg; Universitaire Pers Maastricht, 1995; ISBN 90-5278-188-5; Chapter 8 Problem elicitation in AS and fibromyalgia; accepted for publishing in the Journal of Rheumatology.

Choate R, Smith A, Cardillo J, Thompson L. Training in the use of goal attainment scaling. *Community Mental Health Journal* 1981;17(2):171-181.

EuroQol Group. EuroQoL - a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.

Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. *Clin Exp Dermatol.* 1994 May; 19(3): 210-6.

Hopkins A, ed. Measures of the quality of life and the uses to which such measures may be put. RCP Publications, 1992.

Kaplan RM. Quality of life assessment for cost/utility studies in cancer. *Cancer Treat Rev* 1993;19:85-96.

Kaplan R, Feeny D, Revicki D. Methods for assessing relative importance in preference based outcome measures. *Quality of Life Research* 1993;2:467-475.

Keeney RL, Raiffa H. *Decisions with multiple objectives: preferences and value trade-offs*. New York: Wiley 1976

Kiresuk TJ, Sherman RE Goal attainment scaling: A general method for evaluating community mental health programs. *Community Mental Health Journal*, 4, 1968, 443-453.

Kiresuk TJ, Smith AS, Cardillo JE *Goal Attainment Scaling: Applications, theory and measurement*. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey 1994.

Kiresuk T, Stelmachers Z, Schultz S. Quality assurance and goal attainment scaling. *Professional Psychology* 1982;13(1):145-152.

Krabbe P, Essink-Bot M-L, Bonsel G. On the equivalence of collectively and individually collected responses: Standard-gamble and time-tradeoff Judgments of health states. *Med Decis Making* 1996;16:120-132.

Krahn M, Mahoney J, Eckman M, Trachtenberd J, Pauker S, Detsky A. Screening for prostate cancer. A decision analytic view. *JAMA* 1994;272(10):773-780.

Krumins P, Fihn S, Kent D. Symptom severity and patient's values in the decision to perform a transurethral resection of the prostate. *Med Decis Making* 1988;8:1-8.

Law M, Baptiste S, McColl MA, Opzoomer A, Polatajko H, Pollock N The Canadian occupational performance measure: An outcome measure for occupational therapy. *CJOT*, 57, 1990, 82-87.

Liss P-E *Health care need. Meaning and measurement*. Linköping Studies in Arts and Science 53. Linköping 1990.

O'Leary J, Fairclough D, Jankowski M, Weeks J Comparison of Time-tradeoff utilities and rating scale values of cancer patients and their relatives: evidence for a possible plateau relationship. *Med Decis Making* 1995;15:132-137.

Lorentsen O, Hem K-G *Critical factors and general outcomes of assistive technology*. EU TIDE study *Cost-effective rehabilitation technology through appropriate indicators*

(CERTAIN). Commission of the European Communities, Directorate-General DG XIII, November 1996.

Oortwijn W, van Beekum T Validation of the evaluation method - Report on course and agreements with other projects. EU TIDE study Cost-effective rehabilitation technology through appropriate indicators (CERTAIN). Commission of the European Communities, Directorate-General DG XIII, January 1996a.

Oortwijn W, van Beekum T Experiences from validation. EU TIDE study Cost-effective rehabilitation technology through appropriate indicators (CERTAIN). Commission of the European Communities, Directorate-General DG XIII, October 1996b.

Persson J, Brodin H Legal and macroeconomic factors impacting rehabilitation technology availability. D.2 Chapter 1: Existing socio-economic models. EC TIDE study Horizontal European Activities in Rehabilitation Technology. Commission of the European Communities, Directorate-General DG XIII, August 1993.

Persson J, Brodin H Prototype tool for assistive technology cost and utility evaluation. Deliverable 2, EU TIDE study Cost-effective rehabilitation technology through appropriate indicators (CERTAIN). Commission of the European Communities, Directorate-General DG XIII. CMT report 1996:12, 1996.

Persson J, Brodin H, Lorentsen O, Hem K-G, Andrich R, Ferrario M, van Beekum T, Oortwijn W Cost-Effective Rehabilitation Technology through Appropriate Indicators - CERTAIN TIDE Horizontal Activity TP1264: Final Report, Part A. Public Report. March 22, 1997.

Samuelsson K, Wressle E Kvalitetsutveckling - Dokumentation av arbetsterapeutiska insatser. Memo, Neurocentrum och Geriatriska kliniken, Universitetssjukhuset, Linköping 1996 (in Swedish).

Spreadbury P The binary individualised outcome measure: A case study from occupational therapy. Background paper, ECHHO-SPRI conference "Outcome measures make sense, do they make a difference?", Linköping June 12-13, 1997.

Stalmeier P, Bezembinder T, Unic I. Proportional heuristics in time tradeoff and conjoint measurement. *Med Decis Making* 1996;16:36-44.

Tsevat J, Cook E, Green M, et al. Health values of the seriously ill. *Ann Intern Med* 1994;122:514-520.

Tsevat J, Goldman L, Soukup J, et al. Stability of time-tradeoff utilities in survivors of myocardial infarction. *Med Decis Making* 1993;13:161-165.

Walker S, Rosser R, eds. *Quality of life assessments: Key issues in the 1990s*. Kluwer Academic Publishers, 1993.