

# A Multimedia Preference-Assessment Tool for Functional Outcomes

Mary Kane Goldstein, M.D., David Michelson, B.A., Ann E. Clarke, Leslie A. Lenert, M.D.  
Department of Medicine, Stanford University

## ABSTRACT

*Functional outcomes of clinical trials are often reported as number of dependencies in activities of daily living (ADLs.) Quality-weighting for the ADLs has not been reported. We designed and pilot-tested ADLIB (ADL Index Builder), a multimedia computer program, that presents ADL health states to subjects and elicits from subjects a rating for the quality of life of each health state. Subjects, who were patients over age 50 without previous computer experience, found the program easy to use. Health care professionals specializing in geriatrics confirmed that the ADL presentations used in the program are in accord with typical practice in scoring ADLs. We plan to use the program to obtain population-based preference ratings that can be used to assess efficacy of clinical trials and to provide quality-weights for cost-effectiveness analysis.*

## INTRODUCTION

We describe here the design and pilot testing of ADLIB, a multimedia computer program that presents ADL health states to subjects and elicits from the subjects a rating for the quality of life of each health state.

## BACKGROUND

Cost-effectiveness or cost-utility analysis (CUA) provides a method for comparison of alternative treatments or programs by health-policy decision makers [1]. CUA requires a measure of health effect that summarizes outcomes of diverse conditions in a single measure. Quality-adjusted life years (QALYs), which aggregate in a single measure the total health improvement for a group of individuals, provide such a summary measure. Quality-adjustment weights for the QALYs are obtained by elicitation of preference ratings from a group of patients, from members of the general public, or from some other group whose values are deemed important by the decision-makers [2]. In the past, researchers have obtained preference ratings from the general public by having large numbers of individuals rate the quality of life of various health states as presented in standardized written descriptions [3-5].

Much of medical care of adults—particularly care of older adults who are the largest users of the health-care system—now focuses on improvements in functional status, rather than on cure of disease. One of the most widely used of the many functional status health measurements is the Activities of Daily Living (ADL) scale. This scale includes basic

activities of self-care, such as dressing, and feeding. Results with the scale are typically reported as the number of different ADLs with which an individual requires assistance, referred to as number of dependencies. Although there are large databases of ADL dependencies for different populations, and the ADLs have been used as outcome measures in scores of clinical trials, little is known about how patients or the general public value changes in one ADL relative to another. To perform CUA of health-care interventions that have a major effect on functional status, we must obtain preference ratings for ADLs.

The process of obtaining preference ratings proceeds as follows [6]. Researchers define a health-state classification system pertinent to the problem at hand. Then, they choose representative scenarios for each health state and present a description of these states to subjects, who rate the quality of life in each health state. The task faced by the subjects who perform the rating consists of two sub-tasks. The first is a learning task, in which the subject must come to understand the health-state description in the way intended by the researchers. The second is a rating task, in which the subject must use a standard method for designating the quality of life in the health state. Health-state descriptions have typically been presented in the form of a written description. In a few cases, the written descriptions have been supplemented by other material, such as an audiotape demonstrating voice quality [7] or cartoons depicting level of disability [8]. Multimedia presentations can enhance the subject's understanding of the health state to be rated [9].

## DESIGN CONSIDERATIONS

### Functional Outcomes

The oldest and most widely used ADL scale is the Katz Index [10]. This scale includes bathing, dressing, toileting, continence, feeding, and transferring. Another well-known scale is the Lawton Personal Self-Maintenance Scale, which uses a different method of scoring each ADL, and which adds grooming and ambulation. Many other health-status measurement scales include one or more of the ADLs as part of a measure of functional status—for example, the Health Assessment Scale (HAQ) [11], and the measures from the Medical Outcome Study [12]. The Quality of Well-Being (QWB) Scale [13] provides quality ratings, derived from presentation of written health-state descriptions to members of the general public, for a variety of health states with differing functional status; however, the QWB does not provide fine enough gradations to yield

preference ratings that distinguish among ADLs. We wanted to obtain preference ratings for health states that would correspond closely to health-state data obtained in large clinical trials, so that our results could be used as estimates of the quality weighting factors in CUA of such trials. We chose the Katz scale as the most widely used. We supplemented the Katz categories with grooming and ambulation from the Lawton scale, so that our results could also be applied to trials that use that scale. The Katz scale categories of full dependence in eating and in continence include both the need for human assistance and the need for a catheter. Based on discussions with patients at a geriatric clinic, we believed that patients were likely to have ratings for catheter dependence quite different from those for need for human assistance, so we asked for separate ratings for catheter use.

### Program Design Goals

We anticipated using the program to obtain ratings from individuals who are at risk for, or who already have, conditions that cause ADL dependencies. The Department of Veterans Affairs (VA) is planning two large multicenter trials of a geriatric intervention in which ADLs are an important outcome variable, so we wanted to be able to use the program with older adult VA hospital patients. We set the following goals for the design of the computer program: (1) the program must be easy to use for persons who had no computer experience; (2) the content must be clear and comprehensible even for subjects who have only grade-school level comprehension; (3) the program must provide enough visual and auditory material to maintain interest and focused attention, but must not overwhelm the subject with information; (4) to minimize the risk of interviewer bias, the program must include standardized explanation of the health states and of the rating task; (5) the concept of dependency in the ADLs must be presented in a balanced way that neither conceals the level of disability nor interjects assumptions about emotional state; (6) the material presented must not be offensive to subjects; (7) the rating scale must be easy to use; (8) the whole program must not take more than 30 to 45 minutes on average to run; (9) the program must be able to run on a computer compact and light enough to use in patient rooms in the hospital and to bring to the homes of home-care patients; (10) the trackball must be easy for a person with arthritis to manipulate.

## IMPLEMENTATION OF ADLIB

### Overview of the Program

We used Macintosh computers to develop and present multimedia materials. Each frame presented on the computer screen includes pictures and text, and is accompanied by a voiceover. Sound is used to indicate

the transition from one card to the next. The overall sequence is as follows: introduction to use of the computer and large trackball, rating of current health, presentation of the ADL descriptions and self-categorization of independence or dependency, rating of the quality of life with dependency in one ADL at a time, comparison of the individual ADL ratings for fine-tuning of ratings, rating of combinations of ADL dependencies, generation of report. The program uses the principle of iconification of both pictures and voices. We edited pictures with Adobe Photoshop, recorded and edited sounds with Paracomp SoundEdit, and then integrated them into SuperCard. The presentation was developed using a Mac II Ci computer with a color monitor. We transferred the program to a Powerbook 180 for testing and use with subjects. We used large type and high contrast to maximize visibility for older subjects with low-vision problems.

### Construction of the Program

**Pictorial Material.** We recruited seven VA patients older than 70 years and two members of the nursing staff who were interested in serving as actors for this program. We obtained written permission from each patient (in one case, from the patient's guardian). We took photographs of patients performing ADLs independently or with assistance with a Canon still video camera. The pictures were captured by the Supermac Video Spigot card and ScreenPlay software. The pictures of toileting and bathing provided context to identify the activity (a toilet room with clearly visible toilet-paper roll, a shower chair, and a hand-held shower head), but we respected the patients' modesty by keeping the patients dressed from the waist down. We were careful to show only one impairment at a time in a picture. For example, the patient with a leg bag for urine drained by indwelling foley catheter was shown standing independently, rather than using a wheelchair, and the patient receiving a gastrostomy tube feeding is shown sitting up fully dressed.

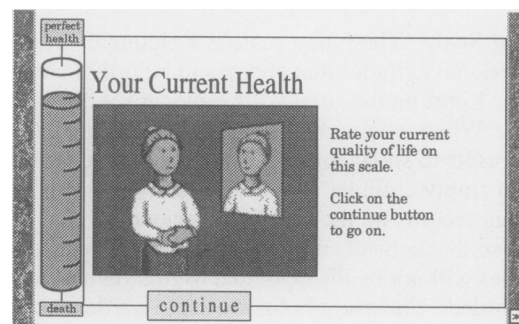


Fig. 1. Rating Quality-of-Life in Current Health

The instructional cards use drawings (Fig. 1); the ADL cards use photographs (Fig. 2.) The people in the pictures represent a mix of races and genders.

**Health-State Descriptions.** We took standard descriptions of independence dependence in the Katz ADLs from published material [14], and re-wrote them to sixth-grade comprehension level. We modified descriptions of grooming and ambulation from the Lawton scale [14]. We recorded MKG reading aloud the final scripts for the ADL descriptions. While the voiceover gives the complete description of the ADL, a picture or several pictures of the ADL activity are shown. Large boxes on the screen hold the words "Independent" and "Dependent," accompanied by several words of description that summarize the voiceover.

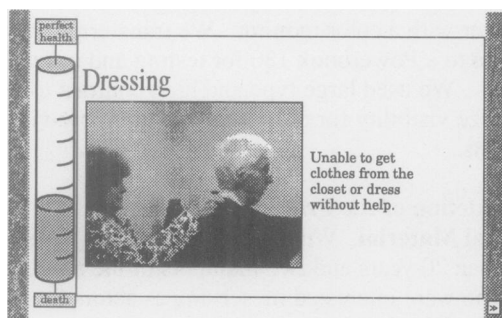


Fig. 2. Rating Quality-of-Life with Dependency in Dressing

**Subject Self-Categorization of ADL Function.**

After each description of an ADL, the program asks the subject to rate himself as independent or dependent in that ADL by clicking on the appropriate box. The self-categorization serves a dual purpose: It provides data on the subject's ADL status, and it engages the subject in an interactive task to enhance memory of the ADL description. Subjects who have impairments in ADLs are presented a separate portion of the program that asks them to rate the quality of life they believe they would have if all elements of their life remained the same EXCEPT that they had a restoration of full function in that ADL.

**Rating Scale.** The rating scale is a picture of a three-dimensional cylinder that appears to be full of colored liquid. Lines on the side of the cylinder set off 10 levels. When a point on the cylinder is selected, the scale resets to show colored liquid below that level and an empty cylinder above. The box in which clicking records on the scale is set widely around the scale, so that subjects who click on the wall of the cylinder will not be disappointed to find their value unrecorded. The anchors for the scale are described as "best possible quality of life" and "worst possible quality of life, or death." Text reminders of the anchors appear on the screen with the rating scale.

**Rating of Quality of Life of the Health State.** The transition from the section of the program that de-

scribes the ADLs to the section with the rating task is marked by a change in sound and by a special card which describes the rating task. The program gives a list of items to consider as subjects make their rating decision. Subjects are reminded several times to imagine dependency in just one ADL at a time.

**Comparison of Individual ADL Ratings.** After rating of each ADL is completed, the subjects are shown the ADLs in comparison, four at a time, in order of quality-of-life rating, from highest to lowest. The ADL picture used on the original card is reproduced in a small version, with a line connecting it to the rating scale to show the level at which dependency in that ADL was rated. Subjects can change their ratings at this point by dragging the picture up or down. When they click the "continue" button, the highest rated ADL slides off the screen to the left and the next ADL slides on. Subjects may use "go back" and "continue" buttons to adjust all the ratings. They indicate satisfaction with the comparative ratings by using a "finished" button.

**Combinations of ADL Dependency.** It is not known whether ADL preference ratings exhibit preference independence (combine additively [7]), so we will need to obtain ratings of combinations of ADL dependencies. Rating all the  $2^8$  possible combinations of eight ADLs at two levels (independent or dependent) would be too arduous a task for subjects, so we selected a subset of possible combinations. Katz described six groupings of ADL impairments that were the most common in his subjects [14]. The Katz groupings account for approximately two-thirds of the groupings found in actual patients. The Katz groupings list certain ADLs "plus one more ADL dependency." To set upper and lower limits on the possible rating of each combination, we test the Katz-defined group plus first the highest rated of the remaining ADLs, and then the lowest rated. For example, one Katz category is "bathing, dressing, plus one more." To obtain the upper limit for this category, we test, for each subject, the combination of dependency in bathing, dressing, and the subject's highest-rated ADL; to obtain the lower limit for this category we test the combination of dependency in bathing, dressing, and the subject's lowest rated ADL. The maximum number of ADL dependencies tested in combination is six.

**Instructions for Moving Around the Program.**

Instructions for self-categorization of ADL function ("How much help do you need with this ADL?") were recorded in a different voice (DJM's) from the one used to give the descriptions of health states. Instructions follow each ADL description and provide a recognizable, repeated, pattern for the self-categorization task. Similarly, the instructions for the rating

tasks include a line that precedes the description of the ADL dependency ("Remember that you are dependent in just this one ADL."), and a line that follows the description ("The scale shows how you rated your current health. How much worse would you rate your quality of life if you were dependent in this ADL?"). These instructions frame the description of the ADL and provide a standardized context for the rating task. Instructions elsewhere in the program—for example, the instructions about the comparisons of all the previously rated ADLs—are recorded in the same voice. Subjects who report dependencies in ADLs receive instructions that are worded differently, to take into account their dependencies. For example, instead of being asked how much lower their quality of life would be if they were dependent in one particular ADL, they are asked how much greater their quality of life would be if they were dependent in ONLY that particular ADL.

### PROGRAM STATUS

During development, portions of the program were discussed with or shown to several health-care professionals and health-services researchers. When a working prototype was completed, we pilot tested the program in two phases: (1) a process of iterative pilot testing and program changes, with patients over age 50 at a VA clinic as subjects, and (2) testing of a single version of the program, with health-care professionals who specialize in geriatrics as subjects. Our aim in testing the program with patients was to assess the feasibility of using this program with the target population—VA patients. Our aim in testing the program with professionals was twofold. We wanted to ensure that the quality-of-life ratings that we obtain with the program are based on descriptions of ADL dependency as typically scored by health professionals when they evaluate patients. We also believed that professionals would be able to assess the suitability of the program in light of experience with a large number of older patients, and that the professionals would be less reticent to voice criticism of the program than patients would be.

**Pilot-test with Patients as Subjects.** Each of six patients went through the program and then answered questions about it in a structured interview. We observed closely and took notes on subjects while they were using the program. We tape-recorded the interview for two subjects, and the complete interaction including the running of the program for three others.

None of these six subjects had ever used a computer before. Each expressed initial hesitation, but willingness to try it. By completion of the session, each subject expressed satisfaction with having "learned how to use a computer." Questions and comments from

the subjects led us to make many program modifications, including rewording of the anchors on the screen, rewording of the health-state descriptions, clarification of the instructions in the rating task, insertion of a transition card between sections of the program, clarification of the meaning of "quality of life" as used in this program, modification of the descriptions to underscore the distinction between toileting and continence, and substitution of different pictures for dark or unclear pictures. Although the instructions for the rating task had seemed clear to us and to those of our colleagues who had previously viewed the program, we found that two subjects—both of whom were fully independent in all ADLs—rated dependence in one ADL at a level higher than the rating they gave their current quality of life. Discussion with them of their understanding of the task enabled us to rewrite the instructions to help users to understand the cognitive leap from rating current health to rating a hypothetical state of dependence in an ADL.

These subjects all said, when asked, that the descriptions of the ADLs were clear and easy to understand, that the trackball was easy to use after the initial practice, that there was nothing offensive or disturbing in the program, that the text was easy to see, and that the process of rating on the scale made sense.

**Pilot-Test with Health Care Professionals.** Ten health-care professionals (six female and four male) completed this phase of pilot-testing. The subject first completed the program (approximately 30 minutes), then filled out a written questionnaire, and then participated in a structured interview about the program. The questionnaire asked subjects to write down everything they recalled about the presentations on dressing, toileting and continence, as a check on the content that had been communicated to the subject [10]. Next, subjects were asked, for each ADL, whether they thought the presentation gave an accurate description. Then, subjects were asked to circle a number from 1 to 5 indicating how easy or hard it had been to do the rating task (1 = very easy). In the interview, subjects were asked whether the sound and pictures in the presentation were clear, whether the presentation was sufficiently easy for cognitively intact frail geriatric patients to understand, and whether anything in the presentation would be offensive to patients. Subjects were also asked to say, for each Katz ADL in turn, whether they could imagine a realistic scenario in which a patient had a dependency in just that one ADL. Responses to this question were probed for further detail.

**Results of pilot test with professionals.** Subjects were observed to attend closely to the entire presentation. All subjects accurately recalled the essential facts of

the three ADL descriptions, including the particular distinction between toileting and continence made in the Katz scale. Eight subjects felt that all eight ADLs had been presented accurately. The mean response for how difficult it was to do the rating was 3.2, range 1 to 5. All subjects thought the program was easy enough for patients to understand, and that the sound was clear. Seven subjects said that the pictures were clear and easy to identify, two said that the small pictures used in the comparison section were hard to see ("but you could tell what they were from the previous presentation"). No subject thought the presentation would be offensive to patients. All subjects agreed that it is unrealistic to posit an isolated dependency in transferring, and seven thought this of toileting. All agreed that isolated dependency in continence is realistic. There was substantial disagreement about whether it is realistic to present isolated dependency in bathing, dressing, and eating. The most common comments were a positive reaction to the program overall, and a statement of appreciation for the opportunity to do comparisons after the individual rankings.

#### DISCUSSION AND FUTURE PLANS

We have developed and performed initial testing of ADLIB, a multimedia computer program that elicits preference ratings for ADLs from subjects. Subjects attended closely to the program, commented positively on it, and performed well on a recall task about program content. Professionals in the field of geriatrics found the program suitable for use with geriatric patients.

Based on the pilot testing, we are revising the program to assess toileting and transferring dependencies only as part of combinations of ADL dependency. We are also revising and retesting the descriptions of bathing, dressing, and eating to ensure that patients can imagine an isolated dependency in each.

After completing the changes indicated by the pilot-testing, we shall use the program with a large number of older subjects to obtain population-based preference ratings for functional outcomes defined by the ADLs. We hope to obtain ratings that can be applied to ADL outcome information from clinical trials to provide a more complete picture of the efficacy of clinical interventions, and that can be used as quality weighting factors to calculate QALYs for CUA.

#### ACKNOWLEDGEMENTS

This work is supported by an NRSA fellowship funded by the Agency for Health Care Policy and Research (MKG), the Hartford Foundation Stanford Geriatric Center of Excellence (AEC, LAL), the Palo

Alto VAMC HSR&D (DJM), Fonds de la Recherche en Sante du Quebec (AEC), the Pharmaceutical Manufacturer's Association (LAL), and the Stanford University CAMIS project funded by NLM.

#### References

1. Kamlet MS. A Framework for Cost-Utility Analysis of Government Health Care Programs. In: Office of Disease Prevention and Health Promotion, Public Health Service, U.S. Dept of HHS, 1992.
2. Torrance GW, Feeny D. Utilities and Quality-Adjusted Life Years. *Int J Technol Health Care* 1989;5:559-575.
3. Kaplan RM, Anderson JP. A general health policy model: update and applications. *Health Serv Res* 1988;23(2):203-235.
4. Cadman D, Goldsmith C. Construction of Social Value of Utility-Based Health Indices: The usefulness of factorial experimental design plans. *J Chronic Dis* 1986;39(8):643-651.
5. Torrance G. Measurement of Health State Utilities for Economic Appraisal. *J Health Econ* 1986;5:1-30.
6. Froberg DG, Kane RL. Methodology for Measuring Health-State Preferences - I: Measurement Strategies. *J Clin Epi* 1989;42(4):345-354.
7. McNeil BJ, Weichselbaum R, Pauker SG. Speech and Survival: Tradeoffs between Quality and Quantity of Life in Laryngeal Cancer. *New Engl J Med* 1981;305(17):982-987.
8. Hadorn DC, Hays RD, Uebersax J, Hauber T. Improving task comprehension in the measurement of health state preferences. A trial of informational cartoon figures and a paired-comparison task. *J Clin Epi* 1992;45(3):233-243.
9. Goldstein MK, Clarke AE, Michelson D, Bergen MR, Lenert LA. Developing and Testing Multimedia-Based Presentations of Health States. Submitted 1993.
10. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function. *J Amer Med Assoc* 1963;185(12):94-99.
11. Fries JF, Spitz P, Kraines RG, Holman HH. Measurement of Patient Outcome in Arthritis. *Arthr and Rheum* 1980;23(2):137-145.
12. Stewart AL, Ware JE, ed. *Measuring Functioning and Well-Being: The Medical Outcomes Study Approach*. Durham, North Carolina: Duke University Press, 1992:449.
13. Kaplan RM, Anderson JP, Wu AW, Mathews WC, Kozin F, Orenstein D. The Quality of Well-being Scale. Applications in AIDS, cystic fibrosis, and arthritis. *Med Care* 1989;S27-43.
14. McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires*. New York, Oxford: Oxford University Press, 1987.