A Catalog of Dermatology Utilities: A Measure of the Burden of Skin Diseases

Suephy C. Chen,*†‡ Ahmed M. Bayoumi,§§§ Seaver L. Soon,* Kent Aftergut,¶ Ponciano Cruz,¶ Shenara A. Sexton,* Calvin O. McCall,* and Mary K. Goldstein,#**††

*Department of Dermatology, Emory University School of Medicine, Atlanta, Georgia, USA; †Emory Center for Outcomes Research, University School of Medicine, Atlanta, Georgia, USA; ‡Department of Health Services Research and Development & Division of Dermatology, Atlanta VA Medical Center, Atlanta, Georgia, USA; §Inner City Health Research Unit, St. Michael's Hospital, Toronto, Ontario, Canada; §§Department of Medicine, University of Toronto, Toronto, Ontario, Canada; ¶Department of Dermatology, UT Southwestern, Dallas, Texas, USA; #VA Palo Alto Health Care System, Palo Alto, California, USA; **Department of Medicine and ††Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California, USA

Utilities are measures of quality of life that reflect the strength of individuals' preferences or values for a particular health outcome. As such, utilities represent a measure of disease burden. The aim of this article is to introduce the concept of utilities to the dermatology community and to present a catalog of dermatology utilities obtained from direct patient interviews. Our data are based on 236 total subjects from Grady Hospital (Atlanta, GA), Stanford Medical Center (Palo Alto, CA), and Parkland Hospital (Dallas, TX). The mean time trade-off utilities ranged from 0.640 for blistering disorders to 1.000 for alopecia, cosmetic, and urticaria. The mean utility across all diagnoses was 0.943. We present utilities for 17 diagnostic categories and discuss the underlying reasons for the significant disease burden that these utilities represent. We also present these dermatology categories relative to noncutaneous diseases to place the cutaneous utilities in perspective and to compare the burden of disease. We have demonstrated that skin diseases have considerable burden of disease and provided a preliminary repository of utility data for future researchers and policy makers.

Key words: quality of life/utility theory/cost-effectiveness analysis
J Investig Dermatol Symp Proc 9:160–168, 2004

In September 2002, the National Institutes of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) sponsored a workshop entitled, "Burden of Skin Disease." Experts in epidemiology, health services research, and quality of life research, as well as representatives from patient advocacy groups, met to discuss the various definitions of "burden," how the definitions of burden pertains to skin diseases, and the importance of measuring such skin disease burden. Quality of life (QOL), economic, societal, and family impact were discussed as aspects of skin disease burden. At the end of the discussion, all participants agreed that the dermatology community should derive core measures of prevalence/incidence, financial costs, and QOL impact of skin diseases to establish a repository of information that defines and establishes the magnitude of skin disease burden. One metric of disease burden, called utilities, was briefly discussed at the workshop and is the subject of this article. We seek to introduce the concept of utilities to the dermatology community and to present a catalog of utilities obtained from direct patient interviews that may serve as a reference for policy makers and researchers.

Utilities are quantitative measures of QOL that reflect the strength of individuals' preferences or values for a particular health outcome, also called "health states" by researchers. Utility scores usually range from 0 to 1, where 0 represents a preference for a health state equal to that for death and 1 represents a preference for a health state equal to that for perfect health. For example, two hypothetical patients may each have a 1-in.-diameter forehead lipoma. The lipomas are identical. Nevertheless, the first patient is a 16-y-old woman with an upcoming high school prom whereas the second patient is a 59-y-old man who cares very little for his appearance. These two patients will have different preferences for the health state of the forehead lipoma. The self-conscious 16-y-old woman may give a utility of 0.76, indicating that she does not want to have the lipoma and considers the lipoma quite a burden. On the other hand, the indifferent 59-y-old man may have a utility of 0.999, indicating that having the lipoma is similar to being in perfect health.

Utilities should be distinguished from the other main type of QOL measure, the health status instrument. Although both metrics attempt to measure the QOL impact of particular health states, they capture different aspects in different ways. Health status instruments describe the nature of health problems by identifying all the important features of a disease state such that every aspect of every dimension is detailed and scored separately (McDowell and Newell, 1996). These instruments measure and score dimensions such as physical, functional, psychological, and social health. The major advantage of the health status instrument lies in the level of detail provided about QOL.

Abbreviations: CEA, cost-effectiveness analyses; QOL, quality of life; TTO, time trade-off

By capturing such detail, health status instruments accurately describe the health status being investigated. These instruments can also detect the specific QOL dimension that changes with a therapeutic intervention.

Although health status instruments can be very comprehensive and descriptive, they face a problem of balancing comprehensiveness and generalizability. There are many disease-specific health status instruments which cover issues related to a specific disease. For example, Scalpdex (Chen *et al*, 2002) is a health status instrument that inquires into issues particular to scalp psoriasis and seborrheic dermatitis. Nevertheless, the results of Scalpdex cannot be compared to results of other disease-specific health status instruments since the issues are different. There are generic health status instruments that can be applied across a variety of diseases, which allow for broad comparisons. The disadvantage of generic instruments is that they often have inadequate focus for specific aspects of QOL and may not detect important issues pertinent to specific diseases. The other disadvantage of health status instruments is that they do not reflect the preferences or values of a given individual or population for a given health state. Because of this, health status instruments are unsuitable to incorporate QOL outcomes into cost-effectiveness analyses. Even if a summary index were provided for a health status instrument, it would be limited to a description of health status, rather than the preference, or values, associated with the health status.

Utilities are derived from economic and decision theory and measured in such a way (see below) that they can be applied and interpreted across diseases and populations. Thus, different health states, both skin- and non-skin-related, can be compared. Utilities focus on the issues pertinent to the specific disease, so unlike generic health status measures, they are sensitive to important issues pertinent to specific diseases. Utilities are also considered by health economists as the most appropriate metric to incorporate QOL impact on health states into cost-effectiveness analyses because they reflect preferences and do not just describe health states. (Gold *et al*, 1996; Drummond *et al*, 1998) They can also be used as outcome measures for clinical trials (Feeny and Torrance, 1989) and treatment regimens (Bennett *et al*, 1991; Nease *et al*, 1995) and in individual clinical decision making (Goldstein and Tsevat, 2000).

Utilities may be measured in several different ways. In general, the interviews are structured such that they ask individuals to make a choice among alternative lotteries on quality and quantity of life (Tolley *et al*, 1994). Two of the most popular methods are standard gamble and time trade-off (TTO). Standard gamble asks the subject to choose between being in their particular health state for the remainder of their life or to take a gamble. The gamble is that they would have a certain probability of returning to perfect health or dying immediately and painlessly (Drummond *et al*, 1998). The TTO method asks the subject to choose between a fixed guaranteed life duration in his or her current health or a shorter duration in the best possible health. The ratio of the time remaining after the trade to the life expectancy of the individual is the TTO-derived utility (Drummond *et al*, 1998). We prefer the TTO method because given that most dermatologic problems are often considered minor health states, we anticipate that the subjects would give more meaningful responses to the TTO, in which they were asked to trade off small amounts of time, than to the standard gamble method, in which they are asked to trade off low probability risks.

Because utilities are measured by asking subjects how much time they would trade or how much risk they would take to not have the disease, utilities are inherently measures of disease burden. Because policy makers are increasingly basing health-care resource allocation and research funding on disease burden, it is imperative that utilities are available for dermatology. For this purpose, we report our utility findings from a larger study. In the larger study, we investigated the feasibility of predicting utilities, using a mathematical model, from a general dermatology health status instrument, Skindex (Chren *et al*, 1997). We were motivated by the fact that there are overlapping constructs between utilities and health status instruments and because utilities are difficult to measure whereas health status measures are easy to administer. Nevertheless, in this article, we will report the catalog of utilities to be used as a preliminary repository of dermatology utility information.

Results

Demographics We recruited a total of 267 subjects, 77 from Grady Hospital, 106 from Stanford Medical Center, and 84 from Parkland Hospital. The proportion of subjects approached who agreed to be interviewed varied by study site: 97% at Grady Hospital, 25% at Stanford Medical Center, and 67% at Parkland Hospital. Language barrier was the primary reason for potential subjects to decline at Parkland Hospital. The Stanford and Grady nonparticipants did not give reasons. Completion rates were 90% at Grady (8 did not complete the TTO portion of the interview), 100% at Stanford, and 98% at Parkland (2 did not complete 8 or more Skindex items). Noncompleters were excluded from the final analysis. As usual for utility studies, we also excluded subjects who did not understand the TTO or who gave inconsistent responses to the utility elicitation questions ($n = 6$). For example, some subjects rated their current health utility including skin problems greater than their skin-only utility ($n = 4$). We also eliminated one subject who scored all the Skindex questions 0.75 or higher but gave a utility score of 0.000, indicating a lack of understanding. Thus, the final data included 250 subjects (68 from Grady, 102 from Stanford, and 80 from Parkland).

Subjects represented a variety of ethnic and demographic groups, although these varied by site (Table I). We found that the differences in age across sites were statistically significant ($p < 0.001$, by ANOVA). Pairwise comparisons revealed that the population at Grady was significantly younger than the populations at the other two sites ($p < 0.001$). We found no significant difference in sex across sites ($p = 0.19$, Pearson's chi-square analysis). There was a significant difference in race across sites ($p < 0.001$), when categorized as "white" and "nonwhite"; each site was significantly different from each other ($p < 0.001$, Pearson's chi-square analysis). Most subjects from Grady were

Table I. Demographics of subjects included in the analysis

Variable	Grady Hospital	Stanford Medical Center	Parkland Hospital	Combined
Number of subjects (% recruited)	68 (88)	102 (96)	80 (95)	250 (94)
Age (years)[a]	36 (12)	50 (18)	49 (15)	46 (17)
Sex[b]				
Male	31 (46)	49 (48)	28 (35)	108 (43)
Female	37 (54)	53 (52)	52 (65)	142 (57)
Race[b]				
Caucasian	5 (7)	83 (81)	35 (44)	123 (49)
African-American	57 (84)	1 (1)	33 (41)	92 (37)
Other (includes Latino and Asian persons)	6 (9)	18 (18)	12 (15)	27 (11)
Education[b]				
Primary school	2 (3)	2 (2)	16 (20)	20 (8)
High school	41 (60)	7 (7)	44 (55)	92 (37)
College	22 (32)	44 (43)	12 (15)	78 (31)
Postgraduate	3 (4)	49 (48)	3 (4)	55 (22)
Income[a]	$18,860 ($18,302)	$90,289 ($58,341)	$12,250 ($13,021)	$44,691 ($53,055)
Number of subjects (%)	68 (100)	95 (93)	80 (100)	243 (97)

[a]Mean (SD).
[b]Number of subjects (%).

African-American whereas those from Stanford were Caucasian; subjects from Parkland were relatively evenly split between these two race groups. We found a significant difference across site in education ($p < 0.001$, Kruskal–Wallis test). Pairwise analyses revealed that each site was significantly different from each other, with Stanford subjects having the most education and Parkland subjects having the least ($p < 0.001$, Wilcoxon rank sum test). We found a significant difference across sites for income ($p < 0.001$, ANOVA). The population at Stanford had a significantly ($p < 0.001$) higher annual income than those at each of the other two sites. These differences were expected, because they reflect the underlying patient populations.

Diagnostic categories We asked both patients and dermatologists to list the diagnosis leading to that visit. Where the patient and the dermatologist differed in the diagnosis, we report the diagnosis given by the dermatologists. We excluded 14 subjects who had multiple dermatologic diagnoses because our purpose is to present a catalog of utilities by diagnostic categories and specific diagnoses. Thus, the final analysis was based on the remaining 236 subjects.

We sorted the diagnoses into 17 broad diagnostic categories (Table II). Many of these categories reflected the morphologic similarities among various specific clinical entities. For example, we classified acne vulgaris, acne rosacea, and milia as "acneiform." The category "neoplasia of uncertain behavior" reflects a health state where the dermatologist biopsied a particular lesion and the patients were uncertain about the diagnosis at the time of the interview. Because we elicited the utility before the subject had any definitive diagnosis, the utility given by the patient reflects the anxiety of an unknown, potentially harmful condition. For four categories, each with only one subject, we could not determine a broader diagnostic category. These were alopecia, cosmetic (subject wanted laser resurfacing), sarcoid, and urticaria.

Several other diagnostic categories also had small numbers of subjects. We present these categories in Table II to demonstrate the wide range in utility. Nevertheless, we will focus our attention on the top six categories (papulosquamous, infection-infestation, dermatitis, acneiform, neoplasm of uncertain behavior, and benign tumor), which collectively represent 73% of the study population.

Utilities The mean TTO utilities ranged from 0.640 for blistering disorders to 1.000 for alopecia, cosmetic, and urticaria (Table II). The mean utility across all diagnoses was 0.943. Of note, we reported the utilities for basal cell and squamous cell carcinomas as a combined category in Table II because many dermatologists and patients group such skin cancers collectively as nonmelanoma skin cancers.

We also analyzed utilities within each of these broad categories for specific diagnoses with five or more subjects (Table III). Diseases with fewer than five subjects that we do not report data include pruritus, lichen planus, pyoderma gangrenosa, insect bites, folliculitis, furuncle, hidradenitis suppurativa, tinea, tinea versicolor, seborrheic dermatitis, stasis dermatitis, pityriasis rosea, chelitis, perioral dermatitis, drug eruption, hand dermatitis, melasma, rule-out squamous cell carcinoma, dermatofibroma, seborrheic keratosis, keloid, basal cell carcinoma, squamous cell carcinoma, discoid lupus, bullous pemphigoid, epidermolysis bullosa acquisita, B cell lymphoma, stasis ulcer, herpes simplex, impetigo, actinic dermatitis, keratosis pilaris,

Table II. Utilities for diagnostic categories

Diagnosis category[a]	Number (% of total)	Mean utility	SD	Median utility
Bullous diseases	2 (1)	0.640	0.198	0.640
Lymphoma	6 (2)	0.820	0.290	1.000
Pruritus and related conditions	9 (4)	0.915	0.145	0.966
Papulosquamous*	13 (5)	0.919	0.114	0.971
Ulcers	4 (2)	0.923	0.154	1.000
Infection-infestation*	35 (14)	0.933	0.189	1.000
Dermatitis*	52 (21)	0.939	0.098	0.986
Acneiform*	30 (12)	0.940	0.120	0.990
Sarcoid	1 (0.4)	0.949	—	0.949
Dyschromia	14 (6)	0.966	0.103	1.000
Neoplasm of uncertain behavior*	35 (14)	0.971	0.047	0.996
Melanoma	2 (1)	0.972	0.039	0.972
Benign tumor*	17 (7)	0.974	0.054	1.000
Nonmelanoma skin cancer	8 (3)	0.976	0.052	1.000
Collagen vascular disease	4 (2)	0.979	0.029	0.989
Cyst	8 (3)	0.980	0.038	1.000
Actinic keratosis	9 (4)	0.981	0.056	1.000
Alopecia	1 (0.4)	0.998	—	0.998
Cosmetic	1 (0.4)	1.000	—	1.000
Urticaria	1 (0.4)	1.000	—	1.000
All diagnoses	236 (100)	0.943	0.124	1.000

[a]Diagnostic categories are listed in order of ascending mean utilities. Categories marked with asterisks comprise 73% of the study population.

Table III. Utilities of specific diagnoses

Category	Diagnosis	Number	Mean utility	SD	Median utility
Lymphoma	Mycosis fungoides	5	0.867	0.298	1.000
Pruritus and related conditions	Prurigo nodularis	5	0.943	0.064	0.966
Papulosquamous	Psoriasis	11	0.907	0.121	0.966
Infection-infestations	Condyloma	5	0.706	0.434	0.958
	Onychomycosis	6	0.988	0.018	0.997
	Warts	7	0.986	0.035	0.999
Dermatitis	Atopic dermatitis	5	0.890	0.134	0.915
	Contact dermatitis	10	0.898	0.159	0.984
	Eczema and xerosis	11	0.968	0.055	1.000
	Lichen simplex chronicus	5	0.987	0.022	1.000
Acneiform	Acne vulgaris	28	0.938	0.124	0.990
Dyschromia	PIH	6	1.000	0.000	1.000
Neoplasm of uncertain behavior	Rule-out basal cell carcinoma	8	0.974	0.040	0.997
	Rule-out nonmelanoma skin cancer	10	0.979	0.036	0.997
	Rule-out melanoma and dysplastic nevi	11	0.979	0.026	0.988
Benign tumor	Acrochordon	7	0.962	0.074	1.000

pityriasis rubra pilaris, Hailey-Hailey disease, acne rosacea, milia, poikiloderma, lipoma, and fibrous papule. We also do not report utilities for four subjects with "neoplasm of uncertain behavior": one with Kaposi's sarcoma, two subjects for whom the differential diagnosis was not provided, and one subject for whom it was unclear if he was informed of his diagnosis before that particular appointment. We encourage future investigators to contact us for unreported data since this catalog represents a work in progress.

Comparison of utilities between sites To ascertain any bias owing to the different interviewers between sites, we compared the overall mean utility scores among the three sites. There was no significant difference ($p = 0.35$, ANOVA). We also compared utility scores across the three sites for several common diagnoses: acne, psoriasis, dermatitis, benign growth, and nonmelanoma skin cancers/actinic keratoses. Again, we did not find significant differences among sites for these diagnoses ($p = 0.79$, 0.5, 0.60, 0.50, and 0.41, respectively).

Discussion

Although other studies have purported to measure utilities in dermatology, only two published articles thus far have followed rigorous utility assessment methods: Zug *et al* (1995) and Lundberg *et al* (1999) elicited time trade-off and standard gamble utilities for psoriasis. Lundberg *et al* (1999) also reported utilities for atopic eczema. In addition to psoriasis and atopic eczema, we have assembled the most comprehensive catalog of dermatologic utilities to date. The data reflect some of the most common dermatoses observed in a general dermatology clinic. Our multisite study design ensured a wide variety of subjects with a diverse socioeconomic and geographic range.

This catalog provides a preliminary repository of utility data for both policy makers and researchers. Policy makers interested in disease burden measures may consider dermatology problems to be minor and have relatively low disease burden. If that were true, we would have found utilities to cluster near 1.000 (perfect health). Instead, we found that although close to half of the utility scores were greater than 0.95, the overall mean utility was 0.943 and the lowest utility was 0.640 for bullous diseases.

To put these utility values in perspective and to compare the burden of skin diseases to noncutaneous diseases, we include a table (Table IV) comparing some of our findings to utilities of nondermatologic health states. Although each study used slightly different elicitation methods, the comparison provides some interesting insights. For example, it is not surprising that the burden of bullous diseases (0.640) were comparable to that of kidney diseases (0.740) or AIDS (0.791) because these conditions are incurable and are associated with a considerable risk of morbidity and mortality. It is also reasonable that benign conditions such as warts (0.986) and acne (0.938) have high utility scores. The relatively high utility for mycoses fungoides (0.867) can be understood if the diseases were of the patch, rather than

Table IV. Time trade-off derived utilities of health states

Health condition	Mean utility	SD	Median utility	TTO method
Prostate cancer, metastatic (Krahn *et al*, 1994)	0.58	NA[a]	NA	"Gambler," an automated graphical utility assessment tool; from a group of 10 physicians, including urologists, radiation oncologists, and internists
Bullous diseases	0.640	0.1975	0.6397	See Materials and Methods
Minor AIDS-defining illness (Bayoumi and Redelmeier, 1999)	0.65	0.33	0.75	Verbal interview, no props
Mild hip osteoarthritis (Laupacis *et al*, 1993)	0.69	0.27	NA	Specific interviewing methods not detailed
Condyloma	0.706	0.4335	0.9578	See Materials and Methods
Kidney disease (Hornberger *et al*, 1992)	0.740	NA	NA	Verbal interview, no props; life expectancy was patient's own value
AIDS (Tsevat *et al*, 1996)	0.790	0.31	NA	Verbal interview, no props, used "ping-pong" method; did not state origin of life expectancy values
Symptomatic HIV (Bayoumi and Redelmeier, 1999)	0.81	0.27	0.96	Verbal interview, no props
Atrial fibrillation (Gage *et al*, 1995)	0.823	0.250	0.923	Computer-based utility elicitation tool, U-titer
Mycosis fungoides	0.867	0.2981	1.0000	See Materials and Methods
HIV without symptoms (Sanders *et al*, 1994)	0.870	0.29	NA	Computer-based utility elicitation tool, U-titer
Breast cancer (Grann *et al*, 1998)	0.89	NA	NA	Verbal interview; questionnaire for health states
Acne vulgaris	0.938	0.124	0.990	See Materials and Methods
Warts	0.986	0.035	0.999	See Materials and Methods

[a]NA, not available.

of the nodular stage. It is harder to understand why the utility for condyloma accuminata (0.706) would be so low, but perhaps the stigmata and recurrent nature of this problem is an explanation. Also, with two of the sites representing county hospitals where many patients delay their medical care, the condyloma seen there may be more advanced and cause more morbidity than otherwise expected. Readers are referred to an excellent reference cataloging 1000 utilities of other diseases (Tengs and Wallace, 2000).

Of the six diagnostic groups composing up 73% of the study population, the utilities range from 0.919 (papulosquamous) to 0.974 (benign tumor). To gain intuition into these values, we take a hypothetical man with 35 y of life expectancy. If the man gives a utility of 0.919 for a papulosquamous process, this would mean that he would be willing to trade 2.8 y of his life to have the remainder of his life without the papulosquamous process. Similarly, he would be willing to trade 0.9 y to not have a benign tumor. From this perspective, the reader could reasonably agree that the skin processes in these main diagnostic groups represent a significant disease burden. It is not difficult to discern reasons underlying the burden.

The papulosquamous category consists mainly of psoriasis (mean utility, 0.907; Table III). It is not surprising to find such a low mean utility for psoriasis given the accompanying pruritus, scaling, inflammation, and potential social stigma (Updike, 1976). The infection-infestation category (0.933) includes condyloma, which has been discussed above, and tinea, which can be quite pruritic. Given the chronic, symptomatic nature of many types of dermatitis (0.939), especially atopic and contact dermatitis, it is again not surprising that these patients indicated a significant burden of disease. The mean utility for the acneiform category (0.940) can be explained by the fact that most acne occurs on the face and may affect self-esteem. The mean utility for the category of neoplasm of uncertain behavior (0.971) reflects the anxiety of an unknown, potentially harmful condition. The category of benign tumor includes dermatofibromas, acrochordons, and seborrheic keratoses, all of which can be bothersome to the patient.

As demonstrated, the catalog of utilities provides insight into the burden of skin disease. The catalog will also be useful as a reference for future cost-effectiveness analyses because empirical elicitation of utilities is time-consuming and labor-intensive. Nevertheless, in using our data, researchers need to be aware of several limitations. The main limitation is that many disease categories had relatively few subjects. Although the standard deviation gives insight to the distribution of utility scores, our data are insufficient to speculate about the characteristics of an entire patient population.

The relative paucity of subjects in certain disease categories may potentially threaten the face validity of our data. Nevertheless, our mean utility (0.907) for psoriasis is comparable to that of Zug *et al* (1995) (mild psoriasis, 0.89; moderate, 0.79; and severe, 0.59) and Lundberg *et al* (1999) (0.88). Our atopic dermatitis utility (0.890) is slightly lower than that of Lundberg's eczema (0.93), but our "eczema and xerosis" category is higher (0.968) and may compensate for the atopic dermatitis utility if Lundberg consolidated both categories.

To account for the variability in data from the small sample sizes, future investigators should use the standard deviation data to derive upper and lower limits of utility for their studies. For example, in cost-effectiveness analyses (CEA), a sensitivity analysis is performed where all variables are varied within reasonable clinical parameters to see if the results of the CEA are sensitive to those variables. Utilities should be varied as described above to see if the CEA is sensitive to utilities. If so, investigators should report that their findings are not robust and may want to consider measuring utilities for their study empirically. We also encourage future investigators to contact us for unreported data as well as updates in our reported data because this catalog represents a work in progress.

Of note, to obtain an estimate of the number of patients required to achieve a given level of accuracy of the mean utilities, we calculated 95% confidence intervals (CI) of varying widths for mean utility. We based these calculations on the mean utility and standard deviation across all diagnoses (0.94 ± 0.07). Those diagnoses with zero variance were not included in the calculations (see Table III). Table V presents the number of subjects required to obtain 95% CI of width 0.1 (e.g., 0.8–0.9), 0.05, 0.02, and 0.01.

Although the average across all utilities is a fair representation of most of the diseases, a few have much larger variances such as condyloma and mycosis fungoides. For condyloma, the number of patients required to achieve a 95% CI of width 0.1 would be 250; for mycosis fungoides the number would be 140.

The second limitation is that we did not characterize disease severity and thus are unable to comment on the impact of disease severity on our measured utilities. For example, there are only two patients in the melanoma category; it is possible that one patient had thin melanoma with a high utility whereas the other had thick melanoma with a low utility, averaging to 0.972, or both patients could have medium-thickness melanoma with the same average utility. For researchers to compare two therapies via CEA, they need the utilities of all possible outcomes of the therapies. Thus, they need the utilities of all different levels of clinical severities. Although we do not provide this information in this catalog, we encourage researchers to interpret our data as references, from which they can extrapolate utilities for different clinical severity states. Because our utility catalog represents empirical data elicited directly from dermatology patients, the data are a better source for utility estimates than extrapolation from nondermatologic diseases.

Table V. Number of subjects required to obtain 95% CI for a mean utility of 0.94

95% CI width	Number
0.10	12
0.05	23
0.02	55
0.01	108

The melanoma example brings to light a potential selection bias in the utilities of cancer patients in our study. There are specialty melanoma clinics at Parkland and Stanford, so subjects seen in the general dermatology clinic either were not yet referred to the specialty clinics or were returning for follow-up after being in long-term remission. Because we approached the subjects before their appointment, new patients may not have realized the seriousness of their condition, and thus their utilities may not reflect these concerns.

Finally, readers should note that by using average population-based life expectancy to calculate the TTO utilities, we may have overestimated the life expectancy for people with severe, life-threatening diseases such as collagen vascular disease, malignancy, pyoderma gangrenosum, and autoimmune blistering diseases. We chose to use population-based life expectancies because we were enrolling consecutive patients where we could not anticipate their disease. With population life tables, we would expect the life expectancies to reflect an average across people with severe diseases and healthy people. Nevertheless, the life expectancy issue raises a more controversial philosophical issue: should the "best possible health" option be presented to individuals with the time frame of their disease or the time frame of the average person in "best possible health"? Currently, there is no consensus on the best approach and we do not feel that there are any *a priori* reasons to think one is better than the other, but encourage future investigation to explore the issue.

Despite these limitations, we are confident that our catalog of utilities represents a useful information repository of skin disease burden. Future work should be encouraged to verify the accuracy of these data with adequate sample size. We also encourage future investigators to document clinical severity for future utility assessment studies.

Materials and Methods

Subjects and setting We recruited subjects from general dermatology clinics at Stanford Medical Center (Stanford, CA), Grady Hospital (Atlanta, GA), and Parkland Hospital (Dallas, TX), representing diverse racial, geographic, and socioeconomic groups. We obtained approval from the institutional review board of each respective institution and informed consent from all subjects. The study adhered to the Declaration of Helsinki Guidelines for human subject research. The interviewers recruited subjects as follows: for Grady and Parkland Hospitals, the interviewers consulted the schedule of appointments at each clinic and called each subject from the waiting room in a sequential manner until the first person agreed to participate. After the 3-min interview, the interviewer recruited the next consenting subject on the roster. At Stanford Medical Center, the 30-min interviews were scheduled by telephone calls in the days preceding the clinic visit.

We excluded subjects who could not converse with the interviewer owing to language, physical, or cognitive barriers. We also excluded subjects who could not see the prop used (Fig 1) for utility elicitation. Subjects who had difficulty understanding the utility elicitation procedure were asked to explain, in their own words, the purpose of the questions. If subjects were unable to

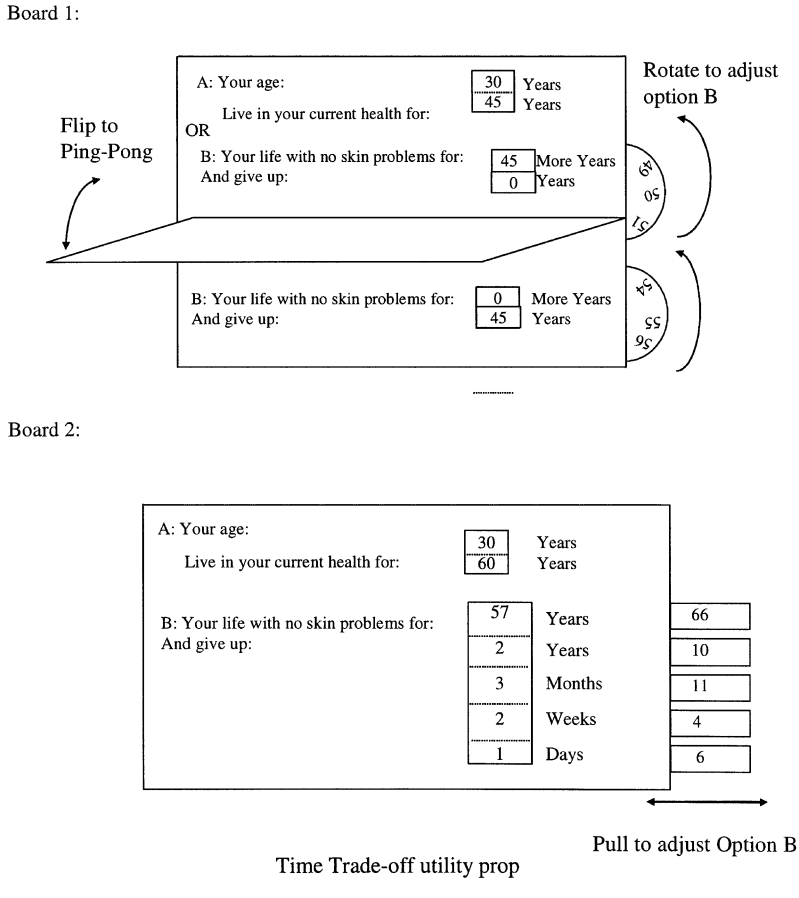

Figure 1
Prop for utility assessment using time trade-off. The prop consisted of two boards. Interviewers showed the board on the top (Board 1) to the subject first, querying her preferences between option A or option B. Ages and corresponding life expectancies were printed on levers that were pulled to display the subject's information through the windows in option A. Option B was controlled using two concentric wheels. The wheels were rotated to adjust the numbers displayed in the windows of option B. A panel was flipped up and down to reveal the decreasing extremes of option B, thereby achieving the ping-pong technique. The interview would continue until options A *and* B were the same to the subject. The interviewers used the board on the bottom (Board 2) to fine tune the subject's time trade-off preference to the level of days. The numbers displayed in the windows were adjusted by pulling on printed levers.

give a satisfactory answer, the interviewer outlined a standard explanation. Subjects who still did not understand the utility elicitation procedure were excluded.

Data collection One interviewer administered all questionnaires at a given site. One of the authors (S.C.C.) trained all three interviewers to ensure interinterviewer consistency. The interviewers collected data on age, sex, ethnicity, skin condition, and comorbid conditions at the beginning of the interview. The subjects next completed the Skindex questionnaire and the utility elicitation, each of which was interviewer-administered. Because income and education are sensitive topics, we reserved these questions for the end of the interview.

Utility elicitation We used the TTO technique for utility elicitation (Stalmeier *et al*, 2001). With this method, subjects are asked to choose between a fixed guaranteed life duration in their current health or a shorter duration in best possible health. We varied the duration spent in best possible health to find the trade-off at which individuals are indifferent between the two choices. The ratio of time remaining after the trade to the guaranteed life duration in current health yields the TTO utility. We obtained the life expectancy from tables of the expectation of life in the United States based on age and sex (Centers for Disease Control and Prevention, 1996). We told subjects that these were average life expectancies for a person of their age. To separate the utility for subjects' skin condition from other health-related problems, we asked subjects to provide separate ratings for the two health states: (1) their current health condition (skin and other comorbidities) and (2) their current skin condition independent from comorbid conditions (Harris and Nease, 1997).

We administered the TTO questions with semiscripted interviews (Appendix I) and physical props (Fig 1). (Dolan *et al*, 1996) The prop consisted of two boards. The first board offered a choice between option (A), remaining life-years in current health, or option (B), a shorter life duration without any skin problems or comorbid conditions. To prevent anchoring effects, the allotted time in (B) was varied in a ping-pong fashion (Dolan *et al*, 1996) until the subject was indifferent between (A) and (B). The ping-pong method involves starting at one extreme (e.g., trading no time) and then presenting the other extreme (e.g., trading the remainder of life) and alternating at the two ends with more/less time traded (Appendix I). The second board fine-tuned the time trade-off into months, weeks, and days. The entire process was then repeated, with the difference that (B) was defined as life without skin problems but with other current health problems. To avoid framing effects (Blumenschein and Johannesson, 1998), the interviewers phrased option (B) to reflect both a loss and a gain: subjects would lose a certain amount of lifetime to gain the remaining lifetime in health without the skin problem. Before the actual utility elicitation, the subjects underwent an example evaluating the health state of paralysis to become familiar with the prop and the technique.

Statistical analysis We used the SAS statistical software package (SAS Institute, 1989–1996) for all analyses. The statistical methods were standard parametric, nonparametric, and chi-square frequency procedures specified in the presentation of the results. Nominal p values are given, without allowance for multiplicity.

This project was supported in part from the NIH/NIA 5 RO1 AG15110, "Disutility of Functional Limitation in the Elderly," an American Skin Association Health Services Research Grant, and an Emory Skin Disease Research Center Pilot and Feasibility grant (No. P30AR42687) from the National Institute on Arthritis and Musculoskeletal and Skin Disease (NIAMS), National Institutes of Health (NIH). S.C.C. is supported in part by a Mentored Patient Oriented Career Development Award (No. K23AR02185–01A1) from NIAMS, NIH and by an American Skin Association. David Martin Carter Research Scholar Award. During the time of data procurement, SCC was funded by a National Research Service Award (T32 HS00028) from the Agency for Healthcare Research and Quality. Views expressed are those of the authors and not necessarily those of the Department of Veterans Affairs.

DOI: 10.1046/j.1087-0024.2003.09112.x

Manuscript received May 29, 2003; revised October 3, 2003; accepted for publication October 9, 2003

Address correspondence to: Suephy C. Chen, Department of Dermatology, 5001 Woodruff Memorial Building, 1639 Pierce Drive, Atlanta, GA 30322. Email: suephy@alum.mit.edu

References

Bayoumi AM, Redelmeier DA: Economic methods for measuring the quality of life associated with HIV infection. Qual Life Res 8:471–480, 1999

Bennett K, Torrance G, Tugwell P: Methodologic challenges in the development of utility measures of health-related quality of life in rheumatoid arthritis. Control Clin Trials 12:118S–128S, 1991

Blumenschein K, Johannesson M: An experimental test of question framing in health state utility assessment. Health Policy 45:187–193, 1998

Centers for Disease Control and Prevention. Vital Statistics of the United States: 1992 Life Tables. Vol. II. Hyattsville (MD): Centers for Disease Control and Prevention, National Center for Health Statistics, 1996

Chen S, Yeung J, Chren MM: Scalpdex: Quality of life survey for scalp dermatoses. Arch Dermatol 138:803–807, 2002

Chren MM, Lasek R, Flocke SA, Zyzanski SJ: Improved discriminative and evaluative capability of a refined version of Skindex, a quality-of-life instrument for patients with skin diseases. Arch Dermatol 133:1433–1440, 1997

Dolan P, Gudex C, Kind P, Williams A: Valuing health states: A comparison of methods. J Health Econ 15:209–231, 1996

Drummond M, O'Brien B, Stoddart GL, Torrance G: Methods for the Economic Evaluation of Health Care Programmes. Toronto: Oxford Medical Publishers, 1998

Feeny D, Torrance G: Incorporating utility-based quality-of-life assessment measures in clinical trials: two examples. Med Care 27:S190, 1989

Gage BF, Cardinalli AB, Owens DK: The effect of stroke and stroke prophylaxis with aspirin or warfarin on quality of life. Arch Intern Med 156:1829–1836, 1995

Gold MR, Siegel RG, Weinstein MC (eds): Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996

Goldstein M, Tsevat J: Applying Utility Assessment at the "Bedside." In: Chapman GB, Sonnenberg FA (eds). Decision Making in Health Care: Theory, Psychology, and Applications. Cambridge: Cambridge University Press, 2000; p 313–313

Grann VR, Panageas KS, Whang W, Antman KH, Neugut AI: Decision analysis of prophylactic mastectomy and oophorectomy in BRCA1-positive or BRCA2-positive patients. J Clin Oncol 16:979–985, 1998

Harris RA, Nease RF: The importance of patient preferences for comorbidities in cost-effectiveness analyses. J Health Econ 16:113–119, 1997

Hornberger JC, Redelmeier DA, Petersen J: Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. J Clin Epidemiol 45:505–512, 1992

Krahn MD, Mahoney JE, Eckman MH, Trachtenberg J, Pauker SG, Detsky AS: Screening for prostate cancer: a decision analytic view. JAMA 272: 773–780, 1994

Laupacis A, Bourne R, Rorabeck C, *et al*: The effect of elective total hip replacement on health-related quality of life. J Bone Joint Surg Am 75:1619–1626, 1993

Lundberg L, Johannesson M, Silverdahl M, Hermansson C, Lindberg M: Quality of life, health-state utilities and willingness to pay in patients with psoriasis and atopic eczema. Br J Dermatol 141:1067–1075, 1999

McDowell I, Newell C: Measuring Health: A Guide to Rating Scales and Questionnaires. New York: Oxford University Press, 1996

Nease RF Jr, Kneeland T, O'Connor GT, *et al*: Variations in patient utilities of outcomes of the management of chronic stable angina: Implications for clinical practice guidelines. Ischemic Heart Disease Patient Outcomes Research Team. JAMA 273:1185, 1995

Sanders G, Owens D, *et al*: A computer-based interview to identify HIV risk behaviors and to assess patient preferences for HIV-related health states. In: Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care, Washington, DC, Philadelphia: Hanley and Belfus, 1994

SAS Institute: SAS System. Cary. (NC). SAS Institute. Inc, 1989–96

Stalmeier PF, Goldstein MK, Holmes AM, *et al*: What should be reported in a methods section on utility assessment? Med Decis Making 21:200–207, 2001

Tengs TO, Wallace A: One thousand health-related quality-of-life estimates. Med Care 38:583–637, 2000

Tolley G, Kenkel D, Fabian R (eds): Valuing Health for Policy: An Economic Approach. Chicago: The University of Chicago Press, 1994

Tsevat J, Solzan JG, Kuntz KM, Ragland J, Currier JS, Sell RL, Weinstein MC: Health values of patients infected with human immunodeficiency virus. Med Care 34:44–57, 1996

Updike J: At War with My Skin from the Journal of a Leper. New Yorker, 1976, p 28–33

Zug KA, Littenberg B, Baughman RD, *et al*: Assessing the preferences of patients with psoriasis. Arch Dermatol 131:561–568, 1995

Appendix I: Example of the Time Trade-off Utility Elicitation In this example, the subject is 30-y-old man with severe penile condyloma. He has an average life expectancy of 45 y. Our interviewers were given a script and instructed to guide the subject as follows:

Interviewer Your first choice is whether you want to live the rest of your life for 45 years with your condyloma or live without your condyloma for 45 years and give up nothing. Which do you prefer or are the choices the same?

Subject I choose option B where I would live 45 more years without my condyloma.

Interviewer Your next choice is whether you want to live the rest of your life for 45 years with your condyloma or live without your condyloma for 0 years and give up 45 years. In other words, you'd die this year without any pain or suffering. Which do you prefer or are the choices the same to you?

Subject I do not want to die now. I choose option A, to live 45 more years with my condyloma.

Interviewer Your next choice is whether you want to live the rest of your life for 45 years with your condyloma or live without your condyloma for 44 years and give up 1 years of your life. Which do you prefer or are the choices the same to you?

Subject I would want to give up 1 year of life to have 44 years without my condyloma. I choose option B.

Interviewer Your next choice is whether you want to live the rest of your life for 45 years with your condyloma or live without your condyloma for 1 year and give up 44 years of your life. Which do you prefer or are the choices the same to you?

Subject I do not want to give up 44 years of my life. I choose to live 45 more years with my condyloma.

Interviewer Your next choice is whether you want to live the rest of your life for 45 years with your condyloma or live without your condyloma for 43 years and give up 2 years of your life. Which do you prefer or are the choices the same to you?

Subject I do not want to give up 2 years of my life. I choose option A, living 45 years with my condyloma.

Interviewer So this means that you are willing to give up 1 year but not 2 years, right? Now we want to see if we can break this down into months, weeks, or days.

The interviewer would use the second board and continue in the same fashion until the number of years, months, weeks, and days that the subject would be willing to trade was ascertained.