N-1153-HEW

June 1979

APPENDIX C TO SUMMARY REPORT: NEW MEASURES OF DIABETIC PATIENT BEHAVIOR, KNOWLEDGE, AND ATTITUDES

Kent H. Marquis, John E. Ware, Jr., and Daniel A. Relles

A Rand Note

prepared for the

CENTER FOR DISEASE CONTROL, U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Rand
SANTA MONICA, CA. 90406

The research reported herein was performed pursuant to Contract No. 200-77-0722 with the Bureau of Health Education, Center for Disease Control, U.S. Department of Health, Education, and Welfare.

The Rand Publications Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

N-1153-HEW

June 1979

APPENDIX C TO SUMMARY REPORT: NEW MEASURES OF DIABETIC PATIENT BEHAVIOR, KNOWLEDGE, AND ATTITUDES

Kent H. Marquis, John E. Ware, Jr., and Daniel A. Relles

A Rand Note

prepared for the

CENTER FOR DISEASE CONTROL, U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Rand
SANTA MONICA, CA. 90406

PREFACE

To assist in the evaluation of diabetic patient education, the Bureau of Health Education of the Center for Disease Control, U.S. Department of Health, Education, and Welfare, sought to identify and develop reliable measures of patient (a) performance (behavior) in managing the disease, (b) knowledge of treatment tasks, and (c) attitudes toward carrying them out. This research effort was conducted by The Rand Corporation under Contract 200-77-0722.

This is one of three appendix volumes under the contract which support the final Summary Report:

Measures of Diabetic Patient Knowledge, Attitudes, and Behavior Regarding Self-Care: Summary Report, Kent H. Marquis and John E. Ware, June 1979 (R-2480-HEW).

The appendix volumes are as follows:

- *Appendix A to Summary Report: Variables Judged Important in Compliance with Diabetes Treatment Regimens*, Kent H. Marquis and John E. Ware, June 1979 (N-1151-HEW).
- *Appendix B to Summary Report: An Evaluation of Published Measures of Diabetes Self-Care Variables*, Kent H. Marquis et al., June 1979 (N-1152-HEW).
- *Appendix C to Summary Report: New Measures of Diabetic Patient Behavior, Knowledge, and Attitudes*, Kent H. Marquis, John E. Ware, and Daniel A. Relles, June 1979 (N-1153-HEW).

SUMMARY

The research goal was to develop a battery of reliable and valid questions that measure the self-care knowledge, attitudes, and behavior of diabetic patients in content areas specified by experts as most important.

Questionnaires were developed and administered to about 500 diabetic patients selected from four treatment facilities. Additional information to evaluate measurement properties was obtained from retest questionnaires, interviews with people who knew the patient, and patient skill demonstrations.

Questionnaire items were grouped into scales designed to measure the self-care knowledge, attitude, and behavior concepts judged important by experts. Each of the scales was evaluated for reliability and validity.

At this point in the research, some promising knowledge and behavior measures have been identified and some basic information has been obtained that will be useful for developing appropriate measures of attitudes. Problems have been identified that require further research to resolve. More research will be needed to develop additional knowledge, behavior, and attitude scales, to minimize the effects of the compliance reporting bias, to increase reliability and validity to remove the possible nonresponse bias.

CONTENTS

PREFACE..... iii

SUMMARY..... v

Section

I. INTRODUCTION..... 1
Patient Sample..... 1
Data Collection Procedures and Completion Rates..... 2
Analysis Preparation..... 3
Analysis Approaches..... 4
Limitations on the Interpretations..... 6
Limitations on Use..... 7

II. BEHAVIOR MEASURES..... 9
Summary of Findings..... 9
Behavior Index Evaluation: Preliminary Steps and Correlations..... 10
Interpretation of Correlations..... 13
The measure of Compliance Reporting Bias and the Effect of an Accurate Reporting Contract..... 15
Behavior Scale Evaluation Results..... 24

III. KNOWLEDGE..... 92
Summary of Findings..... 92
Methodological Notes..... 93
Results..... 99
Validity..... 116

IV. ATTITUDE MEASUREMENT..... 130
Scale Analysis..... 134
Isolating Potential Sources of Answer Variation..... 137

V. RECOMMENDATIONS FOR FURTHER RESEARCH..... 139

I. INTRODUCTION

Evaluations of the new measures of diabetic patient behavior, knowledge, and attitude are given in this appendix.

A panel of medical experts identified characteristics that were most important to measure. (See Appendix B to Summary Report). Questionnaires were developed to measure each important characteristic. Additional questions and other features were added to the measurement design so that the reliability and validity of the original questions could be evaluated.

Diabetic patient samples were drawn in four endocrinology facilities, three located in Southern California and one in the Midwest. The facilities were picked so that the sample contained a range of patient demographic and diagnosis characteristics. Patients were given a first patient questionnaire and sent a second questionnaire about 30 days after the first one was returned. Information was also obtained from the patient's medical record and from an informant, a nonmedical person such as a family member or friend whom the patient said knew most about what the patient did to take care of his or her diabetes.

The analysis addresses the psychological structure of the knowledge, behavior, and attitude trait domains and evaluates the reliability and validity of the measured traits. Reliability is described by the internal consistency of the answers (alpha) and/or by their retest stability. Validity is addressed by noting the amount of response bias in the answers (invalidity) and by correlations with variables or parameters that a valid measure should correlate with (construct and convergent validity).

PATIENT SAMPLE

The diabetic patient sample was selected from four endocrinology treatment facilities. In three of the facilities, the sample was stratified, random selection of persons from records; in the fourth

where access to records could not be obtained, patients who came to a clinic for outpatient treatment of their diabetes were sampled.

496 patients* provided usable first patient questionnaires and are the basic sample in the research. All said they undertook some form of diet therapy, 420 tested their urine, and 391 were classified as insulin dependent. Almost two-thirds (62 percent) were female and 65 percent had at least a high school education. Most were diagnosed as overt diabetics (rather than asymptomatic) and in either good or fair control.

DATA COLLECTION PROCEDURES AND COMPLETION RATES

Five instruments were used in the research. The core document was the first patient questionnaire that was mailed or handed to the sampled patient initially. Distribution and nonresponse follow-up of the first questionnaire was handled by the three mailout facilities to protect the privacy of their patients. Researchers were allowed to distribute and follow up questionnaires in the fourth facility provided they obtained a signed, informed consent statement from the patient indicating willingness to participate in the research. Slightly over 50 percent of the first patient questionnaires were ultimately completed and returned by patients in the first three facilities, 85 percent were obtained from the fourth clinic.

A second questionnaire (with additional questions about knowledge and behavior) was sent to all patients who sent in a usable first patient form. Mailout and follow-up was conducted by the Rand Survey Research Group† and the 80 percent response rate yielded 397 second questionnaires for later analysis.

---

*In 36 instances, a parent filled out the questionnaire for the sampled diabetic patient who was under 12 years of age.

†Patients who returned a usable first questionnaire also signed a statement releasing information to Rand for research purposes. Thus, Rand was able to conduct the second questionnaire mailout and follow up within patient privacy and informed consent guidelines.

Patient respondents named a nonmedical person who knew most about what the patient did to take care of the diabetes. In most cases the named person, called an informant, was a family member. Telephone interviews were completed with 346 of 376 named informants (90 percent). Over 100 cooperating patients were unable or unwilling to name an informant to be contacted.

Medical facilities abstracted the medical records of sampled patients and furnished these abstracts to the researchers upon receipt of the patient's signed permission to release the information. 400 patient record abstracts were available but few contained complete information.

A fifth source of data was the in-clinic observation form administered to a quota sample of about 75 patients attending clinic sessions at one of two facilities. It recorded patient answers to several knowledge questions and

the results of 2 urine sugar tests,
the results of 1 urine acetone test,
the results of measuring the patient's insulin dose,
whether the patient was carrying a diabetic identification,
whether the patient was carrying a source of glucose (sugar).

These data were used to learn more about the relation of answers to the knowledge questions and the skill levels shown during the demonstrations and to evaluate some of the behavior indexes.

ANALYSIS PREPARATION

Before analysis, the questionnaire answers were converted to numeric codes. Answer codes that were out of established numeric ranges were excluded from further analysis. Questionnaire items were deleted that showed either low amounts of between-subject variation (just about everyone gave the same answer), small numbers of responding patients, or retest correlations close to zero. Items were combined where necessary, so that meaningful scores could be obtained. Finally, items were recoded as needed so that high numeric values

indicated correct or favorable answers and low values represented incorrect or unfavorable responses.

ANALYSIS APPROACHES

Analysis approaches differ somewhat for the three categories of items. Each is guided by a general model of answer structure described below.

Hypothesized Answer Structure

A general model, describing the structure of answers to the questionnaire items, guided the measurement analysis. It hypothesizes that answers reflect 3 kinds of "true" score variance along with response biases and random response errors. The expected true score components were:

1. A general factor (level of knowledge about all parts of the self-care regimen, a general attitude about all self-care requirements, or a general tendency to comply or not comply with all prescribed self-care behavior tasks).
2. Global factors (common knowledge, attitudes, or behavior tendencies concerning broad categories of the self-care regimen such as taking medication, testing urine, or diet).
3. Specific factors (knowledge, attitude, and behavior levels for each specific part of the self-care regimen such as timing the urine test reaction carefully).

Different kinds of answer bias were hypothesized for different kinds of questions. We suspected that answers to questions about knowledge might reflect the patient's tendency to guess when he did not know the answer. For behavior, a compliance bias was hypothesized. It is the tendency to claim doing appropriate things even if the patient did not do them.

The random error term acknowledges the inherent variability in answers to questions that cannot be explained by the variables in the analysis.

The relative size of each kind of factor suggests how questionnaire items might be selected and grouped to form scales. If, for example, there is very little difference between a person's attitude towards each part of his regimen, only one scale reflecting this general attitude might be developed. If, on the other hand, a person has very different attitudes toward each important part of his regimen, then it is desirable to build scales to obtain separate measures of each.

If the item response structure is different from that hypothesized, this needs to be known before scale development and the true underlying structure identified. Items are then grouped, when possible, into scales that measure important parts of the true underlying structure.

If response biases are found, it is desirable to construct scales that minimize the effects of bias on the scale score.

Finally, the size of the random error (often called the inconsistency or unreliability) suggests the length of the scale to be constructed. In general, the more items a scale contains, the more reliable the scale score, especially if answers to the items have the same structure.

Obtained Answer Structure

Before beginning the construction of scales, each of the three categories of items was factor analyzed to determine whether the structural model was a useful one. The results of these analyses are described in the sections that follow. In general, however, evidence for the higher level factors (general and global) was not always found. Therefore, the main scale construction effort was concentrated at the specific factor level: Where possible, a scale to measure each of the important behavior, knowledge, and attitude constructs was constructed and evaluated.

Two major implications follow from this scale construction decision. First, the number of scales constructed is large and each scale contains very few items. Second, because the scales are short, the scale scores are somewhat less reliable and valid than they might be otherwise. In addition, if a very short scale contains one "bad" item, it is less likely that the remaining items (if any) will produce a scale score with minimally acceptable measurement properties. Thus, there are specific concepts for which adequate measurements could not be developed completely in this research. Additional items will need to be developed and tested in subsequent research.

Scale Evaluation

Details of the analytic and recommendation approaches differ somewhat for the three categories of items: Knowledge, behavior, and attitudes. Approaches used, therefore, are explained in the appropriate sections. All approaches are based primarily on correlations of "candidate" scales with other variables to assess reliability and validity. Results, interpretations, and recommendations mentioned should be viewed as preliminary at this stage of the research.

LIMITATIONS ON THE INTERPRETATIONS

This is a beginning research effort. As such it has produced new information about promising measurement approaches and kinds of problems to be encountered. It also has its own limitations. The new knowledge is described in later sections. The limitations of the study, affecting interpretations of the data and use of the new scales, are mentioned first.

Caution should be exercised in generalizing the measurement interpretations to other samples and contexts. The patient sample used in the research is not a national sample, it may not be a representative sample of the population from which it was drawn, and the responses were not obtained in an educational intervention context. The effects of these limitations are unknown and further research will be necessary to address and overcome them.

Although not a national sample, efforts were made to represent a wide range of patient disease, regimen, and demographic characteristics in the research samples. The earlier description of the sample suggests that the diversity sought was obtained. The major limitation on interpretations in this area is that we do not know how measurement characteristics of narrowly defined groups may differ from the sample as a whole (e.g., the less education, the poor, small children, patients who do not take medication). Much larger patient samples will be needed for these kinds of inferences.

The low questionnaire response rates (and occasionally high item nonresponse rates) may represent the greatest threat to unbiased generalization of these research findings. Nonresponse bias affects the generalizability of our conclusions to the degree that nonresponders differ in their potential answer stability and validity. It can be assumed that any applied research using these measures will try to achieve response rates in the 80-90 percent range. We do not yet know whether our measurement conclusions will hold under conditions of higher form and item response rates. Further research is essential to resolving the uncertainties about the potential nonobservation biases.

Finally, the measures have not been evaluated as a means for detecting change in regimen compliance due, for example, to an educational intervention. We find that some of the items contain compliance response bias; it is possible that answer bias will increase as a result of an education program that clearly defines expected right and wrong behavior and attitudes. In this case, changes in bias could be misinterpreted as changes in behavior and attitudes. If these measures are used to evaluate change, a careful assessment of post-intervention measurement properties is essential.

LIMITATIONS ON USE

The measures discussed here are still in the development stage and not quite ready for general use. Item and scoring development are sufficiently advanced so that further measurement research might be combined with some program evaluation research.

Within the short-term future, these measures can be refined and used to compare large groups of patients on the important regimen compliance dimensions. Use will be limited to direct comparison of programs until either norms or absolute criterion standards are developed. Use will be limited to large samples because of the scale reliability and validity properties: The properties of the final measures are likely to be good enough for research and evaluation on programs but unlikely to meet standards necessary for assessment of single individuals. Thus, without a substantial amount of further development, the measures should not be used, for example, to diagnose individual patient compliance difficulties before prescribing a self-care regimen or education program.

II. BEHAVIOR MEASURES

This section evaluates the new measures of diabetic patient behavior based on analyses of the questionnaire responses and informant and demonstration information.

SUMMARY OF FINDINGS

A main conclusion is that individual diabetic patients cannot be classified simply as good or bad over a wide range of self-care activities: Knowing that a patient carries out one part of his regimen well does not help to predict how he will comply with other parts of his self-care routine.

Because one behavior does not predict another, measurement is inefficient. If information is desired about a specific form of behavior, a battery of questions about that specific behavior must be asked. Unlike knowledge and attitude, questions about one kind of behavior reveal little or nothing about other kinds of behavior. If all of the behavioral items rated important by the Delphi panel are to be measured, the measurement instrument will contain a large number of items.

Twenty-six behavior scales were constructed to reflect the concepts rated important by the panel. Only three scales met all measurement requirements and 11 scales showed very serious problems that will require intensive further research to understand and correct. The remaining 12 scales are at neither extreme. Four are too short but it is relatively straightforward to add questions in future research to enhance their stability and validity properties. Four have satisfactory stability and true score variance properties but also contain compliance reporting bias variance; if used in applied research, it will be necessary to control for the compliance bias. Four scales appear to have good stability and low bias characteristics but there is some remaining uncertainty about their validity; they probably contain acceptable true score variance characteristics but this needs to be confirmed in future research.

The analysis and interpretation strategies are described next followed by summary evaluations of each of the new behavior measures.

BEHAVIOR INDEX EVALUATION: PRELIMINARY STEPS AND CORRELATIONS

Groups of questionnaire items designed to measure the kinds of behavior judged important by experts are evaluated here, tentatively, for reliability and validity.

Questionnaire items with little or no variation (standard deviation less than .1), items answered by fewer than 50 patients (except oral medication), and items with no stability (retest or parallel form correlations close to zero) were eliminated.

New variables were created as necessary to score each response as compliant or not compliant. In general, the new variables reflect the comparison of what the patient said he did either to what the medical record said he should do or to what the patient thought he should do. Variables were then rescored so that the direction of good and bad answers was consistent across items.

Each answer was converted to a standard score (Z score) and items combined into scales of the Delphi concepts. Within each concept, three scales were created: A scale of "candidate" items to be evaluated, a retest scale, and an informant scale. The candidate scale, usually containing items from the first questionnaire but occasionally new items on the "retest" questionnaire, is the scale that is evaluated for reliability and validity. The retest scale contains relevant items from the retest questionnaire and is used to estimate the test-retest reliability of the candidate scale. The informant scale contains answers given by a knowledgeable other person about relevant parts of the patient's behavior; it is used to evaluate the validity of the candidate scale. In several cases a scale contains only one questionnaire item.

The basic evaluation analysis rests on intercorrelations of these scales to describe reliability and validity characteristics of the candidate index. Interval metric properties of the resulting scales are assumed for the correlation analysis.

The correlations are in three groups labeled "stability," "invalidity," and "validity". Each is described briefly below. In the next subsection, our approach to interpreting the correlations is explained.

Stability

Two indicators of the answer stability of an index are usually available: The test-retest correlation and the alpha coefficient of internal consistency. The retest correlation provides a description of how much answers change across time. Assuming no change in the true trait and no change in answer biases over time, one minus the retest correlation coefficient suggests how much random response variation is in the measured values. Because the retest scales are sometimes shorter than the candidate scales, the reported retest correlation is likely to underestimate the amount of stable variance in the candidate scales. Large retest correlations are desirable but the exact levels needed depend on how the measures are to be used in research; therefore no absolute criterion levels are suggested here.

The Alpha coefficient* is an indicator of internal consistency of the items that form a candidate index. It is sometimes higher than the retest coefficient because it reflects systematic biases that operate during this particular administration of the questionnaire (e.g., the respondent's transitory mood) but do not operate the same way on other occasions (such as retest). The value of alpha is proportional to the number of items in an index and the size of their covariance. In this research the candidate scale sometimes contains more items than the retest scale, so it is not unusual for the alpha value to exceed the retest value. In making interpretations (see below) stability characteristics are inferred based on the arithmetic average of the alpha and retest coefficient.

---

*Cronbach, Lee J., "Coefficient Alpha and the Internal Structure of Tests," Psychometrica, Vol. 16, 1951, pp.297-333.

Compliance Bias

The correlation of the candidate scale with a measure of compliance desirability response bias (called compliance bias) is reported to indicate how sensitive the index scores are to the tendency of respondents to lie about their behavior in a favorable but untrue direction. (Actually, to reduce effects of random error, the arithmetic average of the Form A and Form B correlations is reported.)

A description and evaluation of the compliance bias measures is given later.

Other Biases

We attempted to measure other potential response biases but do not report the results here. The measures of the other biases are of questionable validity and few, if any, of the candidate behavior indexes appear sensitive to them.

Validity

The indicator of validity is based on the correlation between the candidate index and the relevant informant index. Assuming the patient and informant do not share the same response biases, the informant correlation is an indicator of the amount of true score (presumably valid) answer variation in both the patient and informant responses. Positive correlations are desirable and larger coefficients are better than smaller ones. In theory, however, the informant correlation cannot be higher than the stability coefficient.

Occasionally additional validity information is available by comparing questionnaire responses to demonstration scores. Twenty-four patients filled out the first patient questionnaire and also participated in the skill demonstration phase of this research. Their inclusion in both phases is accidental but they should not be viewed as representative of the overall patient sample. Information about questionnaire measurement validity obtained from this group is interpreted impressionistically.

INTERPRETATION OF CORRELATIONS

Using the assumptions and logic that follow, a tentative decomposition of measured index variation into stable variance, bias variance, and true score variance is given. The goal is to suggest both the absolute and relative sizes of the variance components so that future users have some information with which to design their evaluations or other research.

Each candidate index, made up of the average item Z scores, is assumed to contain true score on the trait, response bias, and random response error. Formally, the candidate scale is represented as $C_j = b\ (Trait_j) + d\ (Bias) + E_j$.

Where C_j = Measured value of candidate scale for Delphi trait j (Var $C_j = 1$).

$Trait_j$ = True value of trait j.

Bias = True value of the compliance reporting bias.

E_j = Random response error.

b,d = Parameters to be estimated.

Without loss of generality, assume that the true traits have unit variance. Then b^2 is the amount of true trait variance in the measured candidate, and d^2 is the amount of bias variance in the candidate measure.

Total stable variance, which is $b^2 + d^2$, can be approximated by the retest correlation or the alpha coefficient. The two estimates may differ for reasons discussed above. The arithmetic average is shown in the tables as an estimate of stable variance.

The descriptive estimate of the amount of compliance bias in the candidate index is obtained by correlating the index with the independent measure of compliance bias and correcting for the unreliability of the bias measure: The independent measure of compliance bias is represented by

$$MCB = h\ (Bias) + E_{mcb},\ (Var\ MCB = 1),$$

where "Bias" is the true bias with unit variance and h^2 is the proportion of the measured variance that is true bias variance. Then,

$$\text{Correlation } (C_j, \text{MCB}) = hd.$$

An estimate of d^2 (the bias variance in the candidate index) is obtained by squaring the above correlation and dividing by h^2, the reliability of the independent measure of compliance bias.

A lower bound descriptive estimate of true trait variance in the candidate index is obtained from the correlation of the index with the relevant informant measure.

Assume the informant measure is

$$I_j = p\ (\text{trait}_j) + E_{I_j},\ (\text{Var } I_j = 1).$$

Whatever biases exist in the informant answers are assumed uncorrelated with the patient biases. The correlation of the candidate and informant index scores is pb.

When $p \doteq b$, the above correlation provides the estimate of the true score variance in the candidate scale, b^2. However, there are at least two reasons to suspect that $p < b$: (1) informant indexes almost always contain fewer items than candidate scales and this typically implies a lower value of the coefficient on true score and (2) informants may, indeed, have less exact knowledge of patient behavior.

The difference between the descriptive estimates of stable variation and true score variation in the tables, then may be due to

- o informant indexes that contain more random error than candidate indexes,
- o less informant knowledge about patient behavior than the patient has,
- o unmeasured bias variance in the candidate measure.

THE MEASURE OF COMPLIANCE REPORTING BIAS AND THE EFFECT OF AN ACCURATE REPORTING CONTRACT

The reliability and validity of the compliance bias measure is discussed next followed by the results of a procedure designed to reduce the bias effects. The compliance bias scales have useful measurement properties but the bias-reducing procedure did not produce the hypothesized effects.

Compliance reporting bias has been identified by previous research as a potential source of unwanted answer variance in self-report measures of adhering to prescribed medical regimens. This research, therefore, developed and used questionnaire scales designed to measure each respondent's tendency to give exaggerated self-reports of compliance behavior.

Item Construction

New items, modeled on "lie" scale and "social desirability" scale approaches, were constructed for this study. All concerned self-care health behavior that is desirable but seldom achieved. Each question provides an opportunity for the patient to say something good about himself that is unlikely to be true. A respondent who scores high on the compliance bias index has continually asserted favorable things about himself. While one or two of the answers might conceivably be true, it is very unlikely that a large number of the claims are all true. Under this assumption, high scores on the compliance bias index may indicate a tendency to lie in favorable directions in other parts of the questionnaire, a response tendency that our literature review suggested might be present in self-report compliance measures.

The modal length of similar scales in other research is somewhere around 30 items. To reduce respondent burden, a ten-item scale from the first questionnaire and an eight-item scale from the retest questionnaire was used. Each item had a true and false answer choice.

Within each questionnaire, the direction of the "compliance bias" answer was balanced across items. It is difficult to create items for which a "false" answer choice indicates compliance bias so occasionally the "double negative" construction had to be used (items AC8 and BC1 are examples). Otherwise the questions have a simple structure of the form:

"I do a good thing" (true scored as compliance bias) or
"I do a bad thing" (false scored as compliance bias)

The items used are shown in Table 1. The A prefix indicates a first questionnaire item; the B prefix indicates the item was included on the retest questionnaire. The answer scored as compliance bias is indicated in parentheses beside each item (T = true; F = false).

Retest Reliability and Internal Consistency

The reliability/consistency of the scales is high enough for use in this type of research although the long scales (8 and 10 items) are necessary to achieve these levels. The average inter-item correlation of the ten-item, A-Form compliance bias scale is .10; the internal consistency coefficient, alpha, has an estimated value of .53. The shorter B-Form alpha coefficient is .39 due, in part, to the low average inter-item correlation of .06. The test-retest reliability is .53* for the entire respondent sample and for major subsamples such as the insulin dependent.

All scale items were used in analyses mentioned in this report. If future researchers wish to construct a more homogeneous and efficient scale, consider using the A Form version and deleting items 2 and 3.

---

*Four items appeared in both scales. The A-Form items and their retest reliabilities were: AC2 .42; AC5 .18; AC6 .50; AC8 .28.

Table 1

COMPLIANCE RESPONSE BIAS SCALE ITEMS

AC1. I always get the right amount of exercise every day (T)
AC2. I get at least 8 hours of sleep every night (T)
AC3. Some days I drink less water than I should (F)
AC4. I always do what my doctor's office or clinic tells me to (T)
AC5. I have made mistakes before in taking care of my health (F)
AC6. There have been times when I have not washed my hands before eating (F)
AC7. I inspect my gums for problems every day (T)
AC8. Some days I do not eat the minimum daily required number of vitamins and minerals (F)
AC9. I have never missed taking medicine when I should (T)
AC10. At times I have eaten food that is not good for me (F)

BC1. I don't always get at least 8 hours of sleep at night (F)
BC2. I wash my hands before every meal and snack (T)
BC3. I never make mistakes in taking care of my health (F)
BC4. I use dental floss on my teeth every day (T)
BC5. I drive short distances instead of walking (F)
BC6. I always eat the minimum daily required number of vitamins and minerals (T)
BC7. I eat potato chips or other high calorie snacks every day (F)
BC8. I eat fresh fruit every day (T)

Validity of the Compliance Response Bias Index

Although few empirical tests are available, existing evidence tends to support the validity of the compliance bias indexes. The items have "face validity," because of the way they were written; the scale scores show criterion validity because they correlate positively with two "validation" items in the questionnaire. An ad hoc analysis provides further validity evidence by revealing that the compliance bias scale scores correlate highest with behavior items having the strongest compliance demand characteristics, i.e., items that ask directly whether (or how often) you have done the right thing.

The planned test of criterion validity was the correlation of the scale scores with two items hypothesized to contain only compliance bias answer variance. The correlations are in the expected direction. Items AG2E and BG2E ask whether a food exchange list was used to plan the amount of lettuce or raw tomatoes the patient ate during the past 7 days. Food exchange lists, based on American Diabetes Association recommendations, categorize lettuce and raw tomatoes as free items; protein, carbohydrate, and fat values do not need to be taken into account in planning the meal. It is assumed, therefore, that "yes" answers to these items reflect a compliance reporting bias. The correlations between the compliance bias scale scores and the lettuce/tomato items are positive and shown in Table 1.

Table 2

CORRELATIONS OF COMPLIANCE BIAS SCALE SCORES AND LETTUCE/TOMATO ITEMS

Lettuce/Tomato Item Score	Compliance Bias Scale Form A	Form B
Form A	.25	.22
Form B	.18	.31

Note: N's range from 282 to 354.

Support for the validity of the compliance bias measures also comes from two ad hoc analyses. The first set indicates that the compliance bias indices correlate highest with questions that, with the benefit of hindsight, seem to be logical correlates. The second set indicates that the compliance bias index scores correlate in a sensible direction with respondent demographic characteristics.

The compliance bias scores correlate highest with certain kinds of question items, those that ask the respondent directly whether he complied rather than asking him to indicate his specific actions. These items ask (a) whether (or how often) the patient has behaved as he should or as he has been told to or (b) how often he has done the right or correct thing.* There are 14 such items on the first questionnaire and 10 on the retest after deleting 4 items with insufficient answer variance. Examples include:

AB2 "Did you eat any meals or snacks later than you should have?

AB15 "...on how many days did you take fewer pills than you were supposed to?"

AD12 "Now think about how and when you tested your urine during the past 30 days. Have you done what your doctor's office or clinic told you to do?"

AE6 "Think about what you did to prevent or take care of insulin reactions in the past 30 days. Have you done what your doctor's office or clinic told you to do?"

Table 3 indicates the proportion of times the correlation between the compliance bias index and the retained "Do as told/should" items was positive and statistically significant. The proportions in

*This is somewhat analogous to asking a knowledge question of the form, "Do you know about X?" rather than asking a question that lets the respondent exhibit his knowledge.

all subject matter categories are very high (.74 over all categories). For comparison, the compliance bias index correlated in a statistically significant, positive direction with only 9 other A-Form items.*

Table 3

PROPORTION OF "DO AS TOLD" ITEMS THAT CORRELATED HIGHLY WITH COMPLIANCE REPORTING BIAS

Item Subject Matter	(a) Number Retained Items Positively Correlated With Compliance Bias	(b) Total Number of Retained Items	a/b
Insulin	2	5	.40
Pills	2	3	.67
Nutrition	9	10	.90
Urine Testing	2	3	.67
Safety	2	2	1.00
Total	17	23	.74

Finally, the bias scores were found to be positively correlated with patient age, negatively correlated with education level, and uncorrelated with sex, characteristics of the regimen, and duration of the disease. We view the positive correlation with age and the negative correlation with formal education level as supporting evidence for validity of the scales, to the extent that older or less educated respondents have more difficulty with self-administered questionnaires.

Some potential threats to validity of the compliance bias measures have been noted, also on an ad hoc basis. One is that the

*7 were nutrition items, 2 were urine testing items. Both parts of this analysis are limited to items that can be scored simply (e.g., without reference to the medical record). There about 100 simply scored items in the A questionnaire.

scales, particularly the A-form version, contain a disproportionate number of items about eating and, hence, the total scale variance may reflect true nutrition compliance behavior. Also, the A-Form scale contains a general compliance item, "I always do what my doctor's office or clinic tells me to do," that may also reflect true behavior rather than a response bias. A third criticism is that the underlying scale construction assumptions are too cynical; it is possible that people who get high bias scores really do all the things they claim so that the scores reflect true behavior rather than a tendency toward the response bias.

An analysis of the sensitivity of the measurement bias conclusions to the inclusion/exclusion of the nutrition and "doctor's instruction" items in the compliance bias indexes was undertaken. None of the conclusions changed when the alternative bias indexes were used.

In sum, the compliance reporting bias scales meet the limited tests of reliability and validity that could be applied. Answers to the individual items reflect the underlying bias trait only slightly so long scales are needed to obtain a reliable index of the trait.

Reducing Compliance Reporting Bias: The Honest Answers Contract

An honest answers contract failed to reduce compliance bias or other response errors in this research. Fortunately, it did not increase refusal rates either.

Building on some promising unpublished work by Charles Cannell and his colleagues at The University of Michigan, this research developed a "Careful, Honest Answers Contract" and mailed it to half of the first questionnaire patient sample. It was hoped that the contract would call the respondent's attention to the need for careful reading of the questions and careful consideration of the answers; thus reducing response error problems in the data.

The contract (see Table 4) promised that the researchers would keep confidential any information supplied by the respondent. In return, the respondent was asked to pledge to give careful, honest answers.

Table 4

CAREFUL, HONEST ANSWERS

IF THIS RESEARCH IS TO BE USEFUL, it is important that you think through each question, then answer it as carefully and honestly as you can.

IT IS VERY IMPORTANT that we find out how people with diabetes actually take care of themselves. If you have made mistakes or forgotten to do what you should do, be honest about it. If you have done something perfectly, be honest about that too.

IN THE CONTRACT BELOW, we pledge to keep your answers confidential. We will not show them to your doctor or to anyone else. In return for our pledge, would you agree to read the questions carefully and answer them as truthfully as you can? If you agree, please sign your name on the line marked "my signature." If you cannot answer carefully and honestly, please do not fill out the questionnaire booklet.

CONTRACT

As director of this research study, I promise that your answers and any other information you give us will remain strictly confidential. Your answers will be used only for research purposes and never shown to anyone outside the research project except as required by law.

Kent H. Marquis
Research Study Director

I understand that my answers must be very accurate in order to be useful. This means I must read each question carefully and answer it as truthfully as I can. I agree to do this.

_____
My signature

Be sure to enclose this Contract with the completed questionnaire when you mail it back. The research staff thanks you very much for your time and honesty.

Results indicate that the contract had no effect on the compliance reporting bias, no effect on other response errors, and no effect on the potential nonresponse bias.

We expected that mean compliance bias index scores would be lower for the contract group than the other part of the sample. However, no important mean differences were observed as shown in Table 5.*

Table 5

MEAN COMPLIANCE RESPONSE BIAS FOR CONTRACT AND NO CONTRACT TREATMENTS

Compliance Bias Index	No Contract	Contract	Mean Difference[a]
First Patient Questionnaire	.40	.40	.00
Retest Questionnaire	.53	.55	-.02

[a]Neither difference is statistically significant.

The behavior scale analyses (reported later in this section) were repeated controlling on the contract treatment. Results (not reported here) tended to be in the hypothesized direction but were neither statistically significant nor practically important. No treatment effect trends were observed when the reanalysis was confined to behavior scales showing the largest amount of response bias (e.g., the nutrition scales).

The possibility that an honest answers contract "trades" a reduction of response error effects for an increase in nonresponse bias was examined and found not to have occurred. The contract instructions asked the sampled person not to fill out the questionnaire if he felt he could not do it carefully and honestly. It is possible, therefore, that using

*Only a few respondents who were sent the contract and who returned the questionnaire failed to sign or return the contract. They are included in the "contract" treatment group for these analyses.

the contract would increase the refusal rates and increase the possibility of a nonresponse bias. The trend was in the expected direction but does not indicate an important effect:

	No Contract Sent	Contract Sent	Difference
Number of First Questionnaires received	251	238	13

We sent the contract to half the first questionnaire sample and just under half of the completed questionnaires we received were from this group.

The honest answers contract treatment, in summary, did not appear to have important effects on total error in the mailed questionnaire procedure. It may, of course, have more favorable effects in the personal interview context.

BEHAVIOR SCALE EVALUATION RESULTS

Exploratory analyses failed to identify ways of organizing behavior questions that were more general than the particular concepts specified by the expert panel. Therefore, scales hypothesized to measure the specific behavior concepts were constructed and evaluated.

Twenty six specific scales were analyzed for measurement stability, compliance reporting bias, and validity. Only 3 exhibit satisfactory measurement properties, 12 show enough strengths to recommend further developmental research, but 11 will require extensive new work.

Stability was estimated from the retest correlation and also from correlations between items on the first patient questionnaire. The size of the compliance bias was estimated using the correlation of the compliance bias index with the behavior scale. A lower bound estimate of valid (true trait) variance was made on the basis of the correlation of the behavior scale score with informant reports about the patient's behavior.

Structure of Compliance Behavior

An unexpected finding was that compliance behavior reported by diabetic patients is very multidimensional. Exploratory analyses (e.g., factor analyses) failed to reveal sensible organizing principles that were more comprehensive than the narrowly defined concepts rated important by the expert panel.

We had planned to adopt a confirmatory analysis approach to estimating the variance components of the candidate behavior measures. The response model underlying the confirmatory procedure, however, assumed dependence (positive correlation) of answers to questions about compliance within broad categories of the self-care regimen (e.g., medication, nutrition, safety, urine testing). Neither principal components factor analysis, rotated factor solutions, nor less model-dependent examinations of answer correlations suggested meaningful ways of combining items for the desired analysis approach. The only global organizing principle found was that items can be classified in terms of the presence or absence of compliance response bias variance in their answers.

These results imply that a model which assumes that a patient complies equally well or poorly within global features of his self-care regimen is too simple. Test construction and measurement scale evaluation must be at a very multidimensional level, which means that few items are available to measure each specific concept and the number of concept scales to be constructed and evaluated is large.

The implications for compliance theory and theory based research are equally demanding. Compliance theories need to be expanded to account for the many different levels of regimen compliance for the same diabetic patient. Research on compliance should address the full range of compliance behavior and its many uncorrelated dimensions; intervention strategies aimed at improving compliance require techniques that can affect the many, apparently unrelated, compliance behaviors.

Summary Evaluation of Candidate Behavior Scales

The twenty-six candidate scales hypothesized to measure the important behavior concepts are evaluated here. A summary of the measurement characteristics of 15 potentially promising measures is in Table 6. The first group of scales exhibits satisfactory measurement properties. The second has stability problems that might be corrected by lengthening them. The third group manifests compliance bias variance problems that need to be addressed. Scales in the last group have acceptable stability and little compliance variance but uncertainty remains about their validity.

The evaluations of each of the twenty-six scales are given in the remainder of this section.

Self-Care Safety Measures (SAFSLF)

The expert panel felt some of the most important patient behavior categories concern self-care of the hypoglycemic episode. Six indexes of this topic were constructed, four appear promising, one needs major redevelopment, and the sixth can be dropped. Three potentially good indexes concern simple preventive measures: Carrying sugar, carrying identification, and informing others how to treat an insulin reaction. All have high stability, two have no compliance bias, and two have reasonably high correlations with informant information. Detailed comments about the first measure are given next as an introduction to the tabular presentation of evaluation results; the tabular presentation described may be found on pg. 34.

Two self-report items, AA12 and AB12, make up the index of whether the patient carries emergency sugar. The item list preceding the tables indicates that AA12 refers to "yesterday" and AB12 uses a 7-day reference period. Answers to each item were converted to standard scores and added to form the scale. The measurement evaluation was carried out on the insulin dependent subsample (analysis group: insulin dependent) because the expert panel indicated that carrying sugar was essential only for diabetics at risk for hypoglycemia. Diabetics dependent on external insulin are the largest subset in this sample at risk for hypoglycemia. The scale score correlated

.71 with the appropriate retest scale. The alpha coefficient is based on the covariance of the original two items on the first questionnaire. The arithmetic average of the retest correlation and the alpha coefficient is used as the interpretation of the percentage of stable variance in the index of carrying sugar.

The candidate index is found to correlate .11 with the compliance bias measure, which leads to the interpretation that two percent of the index variance is compliance bias variance, a very small amount.

The index correlates .55 with an appropriate item or items on the informant telephone questionnaire indicating that at least half of the measured index variance (55 percent) covaries with the informant answers.

The preliminary recommendation, given at the bottom of pg. 34, is to "retain" this scale. Its measurement properties appear satisfactory.

For this index, there is an unusual opportunity to compare patient responses to their actual demonstration scores. Recall that both sets of data are available for up to 24 patients who are not necessarily representative of the overall patient sample. The cross-classified questionnaire answers are shown in Table 7 (validity and bias parameter estimates are not made due to the small and possibly unrepresentative sample). The table indicates there is some, but not perfect, agreement between the questionnaire and demonstration data; between 65 and 77 percent of the cases are in the "agreement" cells. Both cross-classifications suggest a small patient compliance reporting bias because more cases are in the lower-left cell than the upper-right cell. The bias does not appear to be large for question AB12; it is more substantial for question AA12.

The index of carrying something identifying the patient as a diabetic has measurement properties similar to the index of carrying sugar. Demonstration data are also available. The demonstration cross-classifications (Table 8) show agreement over 50 percent of the time but a potentially larger compliance response bias than

Table 6

CHARACTERISTICS OF POTENTIALLY PROMISING MEASURES

OF COMPLIANCE BEHAVIOR

Concept	Number of Items	Estimates Of Proportion Of Measured Variance That Is: Stable	Compliance Bias	Valid
TENTATIVELY ACCEPTABLE SCALES				
58 - Carries emergency glucose	2	.82	.02	.55
99 - Takes insulin at appropriate times	3	.67	.00	.41
37 - Achieves and maintains ideal body weight (± 10%)	4	.77	.02	.73
SCALES NEEDING ADDITIONAL ITEMS				
34 - Changes eating time to counteract effects of exercise	1	.46	.01	.12
59 - Reports severe reaction to MD immediately	1	?	.01	.38
60 - Reports moderate or greater levels of urinary ketones to physician	1	?	.00	.37
26 - Tests urine at correct times in reaction to meals	3	.41	.06	.22
SCALES REQUIRING SPECIAL TREATMENT OF BIAS				
61 - Carries diabetic identification	2	.85	Some	.50
53 - Reports symptoms to MD (consecutive high sugar test results)	1	.77	.11	.47
98 - Tests urine for sugar	3	.79	.13	.57
27 - Records urine test if prescribed	3	.91	.08	.48
SCALES NEEDING CONFIRMATION OF VALIDITY				
62 - Tells family/friends insulin reaction	2	.70	.00	.26
21 - Waits correct time before reading result	1	.69	.03	Probably Acceptable
42 - If consumes alcohol, includes in planning calculations	2	.66	.02	.34
52 - Informs all medical care providers of diabetic condition	1	.79	.00	Not Tested

suggested by the questionnaire analysis. As a result, we have classified this scale in the category needing special attention to bias problems.

Table 7

AGREEMENT BETWEEN QUESTIONNAIRE ANSWERS AND DEMONSTRATION INDICATORS OF CARRYING EMERGENCY GLUCOSE (FREQUENCIES)

AA12--Carry Sugar Yesterday?

Demonstration		Yes	No	Total
Shows Sugar	Yes	10	1	11
Source Being	No	7	5	12
Carried (Demonstration)	Total	17	6	23

AB12--Always Carry Sugar (7 Days)?

Demonstration	Yes	No	Total
Shows Sugar	10	2	12
Source Being	3	7	10
Carried (Demonstration)	13	9	22

Note: Missing data have been excluded.

The index for telling others how to care for an insulin reaction has good stability but apparently low validity. Stability is high despite some wording differences between the original and retest questions. The low validity may be an artifact since the estimate is based partly on the correlation of two dissimilar concepts: The patient's answer that at least one family member knows reaction care procedures and whether the informant knows the care procedures. The informant is not necessarily a family member nor is he/she necessarily the family member whom the patient had in mind when giving a positive answer. Although the patient-informant agreement rates are over 85 percent, the small amount of disagreement and the low between-patient variance combine to severely attenuate the validity correlation. Additional validity research is recommended.

Table 8

AGREEMENT BETWEEN QUESTIONNAIRE AND DEMONSTRATION INDICATORS OF CARRYING IDENTIFICATION

AA10--Carries Diabetic Identification (Yesterday)

Demonstration:		Yes	No	Total
Shows Diabetic	Yes	9	3	12
Identification?	No	7	4	11
	Total	16	7	23

BB8--Carried Identification (7 Days)

Demonstration:	Yes	No	Total
Shows Diabetic	11	1	12
Identification?	4	2	6
	15	3	18

Notes: Missing data are excluded. Due to a processing error, we did not include questionnaire item AB11 on the demonstration analysis file. Results for Item BB8, its retest, have been substituted here.

The single-item index reflecting changes in eating to compensate for exercise can serve as a basis for a longer scale with good measurement properties; the single item should not be used by itself.

The retained item, BE5, is from the retest questionnaire; it is worded differently than the first questionnaire item, referring only to eating changes that accompany nonroutine exercise. Because of the wording changes, the "retest" correlation is attenuated although we cannot estimate by how much. A two or three item index, however, should exhibit satisfactory stability properties.

We failed to develop a good measure of patient-initiated changes in the self-care regimen and Delphi index 36, eats in response to hypoglycemic symptoms, is one example. Both the initial question, AE5, and especially its retest, BE5, yielded little between patient variation. This is not because hypoglycemic reactions are rare but because few patients admit not ingesting some form of sugar when reactions occur. The conclusion that patient answers are biased comes from the cross-classification with informant answers shown in Table 9.

Table 9

AGREEMENT BETWEEN PATIENT AND INFORMANT ABOUT WHETHER PATIENT EATS IN RESPONSE TO HYPOGLYCEMIC SYMPTOMS

Patient	Did Not Take Sugar	Took Sugar	Total
Did Not Take Sugar	5	6	11
Took Sugar	41	142	183
Total	46	148	

There are 41 instances where patient claims of taking sugar were not confirmed by the informant and only six cases in the opposite disagreement cell. The interpretation assumes informants do have the correct information. This assumption can be tested by controlling on whether the informant said he knows how to care for an insulin reaction. Controlling on informant knowledge, however, does not change the conclusions drawn.

Questionnaire item AE6 is a general question about whether the patient complied with all safety (insulin reaction prevention and treatment) requirements. Like other general items, the answers contain compliance response bias variance. We recommend that this and similar general questions be dropped from further measurement development and not be used by subject matter researchers.

SAFETY ITEMS

DELPHI 58

AA12 (Yesterday, or on the last day that you were not sick), If you left your house, did you take something with you with sugar in it?

Yes
No
I did not leave my house

AB12 (During the past 7 days, or the last 7 days that you were not sick), About how much of the time did you have something with you with sugar in it for emergencies?

All of the time
Most of the time
Some of the time
None of the time

DELPHI 61

AA10 (Yesterday, or on the last day that you were not sick), If you left your house, did you carry or wear something that says you are diabetic (diabetic ID)?

Yes
No
I did not leave my house

AB11 (During the past 7 days, or the last 7 days that you were not sick), When you were away from home, how often did you carry or wear something that says you are diabetic (Diabetic ID)?

Every time I was away
Most of the time
Some of the time
None of the time

DELPHI 62

AE2 Do any of your family members know what to do if you ever have an insulin reaction?

Yes
No
I don't know

AE3 Do any of your friends know what to do if you ever have an insulin reaction?

Yes
No
I don't know

DELPHI 34

BE5 During the last 30 days when you exercised more than usual how often did you eat something extra to prevent an insulin reaction?

Every time
Most of the time
Some of the time
None of the time
I have not exercised more than usual

DELPHI 36

AE5 The last time you had an insulin reaction, what did you do to keep it from getting worse?

Write in what you did:
I have never had an insulin reaction

GENERAL

AE6 (In the past 30 days). Think about what you did to prevent or take care of insulin reactions in the past 30 days. Have you done what your doctor's office or clinic told you to do?

Yes, exactly
Yes, most of the time
Yes, some of the time
Yes, a little of the time
No, not at all

SAFSLF - DELPHI 58 - CARRIES EMERGENCY GLUCOSE

RETAINED ITEMS: AA12, AB12
DELETED ITEMS: (None)
ANALYSIS GROUP: Insulin Dependent

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.71	
Alpha	.93	.82
INVALIDITY		
Compliance Bias	.11	.02
VALIDITY		
Informant	.55	.55

PRELIMINARY RECOMMENDATION: Index has desirable measurement properties and can be retained.

SAFSLF - DELPHI 61 - CARRIES DIABETIC IDENTIFICATION

RETAINED ITEMS: AA10, AB11
DELETED ITEMS: (None)
ANALYSIS GROUP: Insulin Dependent

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.74	.85
Alpha	.96	
INVALIDITY		
Compliance Bias	.02*	Some
VALIDITY		
Informant	.50	.50

*Demonstration data suggest the presence of a slightly larger compliance bias than estimated here.

PRELIMINARY RECOMMENDATION: The index appears to have satisfactory measurement properties and can be retained. Future use should include a provision to detect and correct for compliance bias.

SAFSLF - DELPHI 62 - TELLS FAMILY/FRIENDS INSULIN REACTION CARE

RETAINED ITEMS: AE2, AE3
DELETED ITEMS: (None)
ANALYSIS GROUP: Insulin Dependent

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.72	.70
Alpha	.68	
INVALIDITY		
Compliance Bias	-.04	.00
VALIDITY		
Informant	.26	.26

PRELIMINARY RECOMMENDATION: Informant questions are not exactly same as patient questions and may lead to an underestimate of validity. Recommend using existing patient items, pending additional validity studies.

SAFSLF - DELPHI 34 - CHANGES EATING TIME TO COUNTERACT EFFECTS OF EXERCISE

RETAINED ITEMS: BE5

DELETED ITEMS: (None)

ANALYSIS GROUP: Insulin Dependent

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest *	.46	.46
Alpha		
INVALIDITY		
Compliance Bias	-.08	.01
VALIDITY		
Informant	.28	.12

*The "retest" item is the one appearing in the first questionnaire. The retest correlation is an underestimate of stability because of differences in question wording.

PRELIMINARY RECOMMENDATION: The retest item has the measurement characteristics expected of a single item scale. It may be retained and combined with one or two additional items (to be developed and evaluated) for a better index of the Delphi concept.

SAFSLF - DELPHI 36 - EATS IN RESPONSE TO HYPOGLYCEMIC SYMPTOMS

RETAINED ITEMS: AE5 (Coded correct if ate or drank something containing sugar or if glucagon was administered)
DELETED ITEMS: (None)
ANALYSIS GROUP: Insulin Dependent

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	Item Omitted*	?
Alpha		
INVALIDITY		
Compliance Bias	-.02**	Some
VALIDITY		
Informant	.15	.15

*Retest item based on BE6 contained insufficient variation for analysis.

**Estimate is attenuated severely by lack of variation in answers to retained item. See text for discussion of the interpretation of some compliance bias.

PRELIMINARY RECOMMENDATION: New development research is needed to create items with more answer variation, less compliance bias, and more validity.

SAFSLF - GENERAL ITEM

RETAINED ITEMS: AE6
DELETED ITEMS: (None)
ANALYSIS GROUP: Insulin Dependent

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.41	.41
Alpha		
INVALIDITY		
Compliance Bias	.22	.09
VALIDITY		
Informant	No Item	?

PRELIMINARY RECOMMENDATION: The item has unsatisfactory measurement properties, as do other general items written in this way. It should not be used in future research.

Reporting Problems to the Physician

A second important area of precautionary self-care is reporting problems and symptoms to the physician before they develop into major difficulties.

Three indexes have been constructed: One was evaluated completely and shows promise if provisions are made to handle response bias. The other two are single-item indexes; while they cannot be evaluated completely, both have validity correlations in the acceptable range for single items and neither appears to contain much compliance bias variation. It is very likely that adding one or two more questions to each index will produce scales with desirable variance properties.

We did not attempt to measure everything implied by Delphi concept 53 (contacts physician for new--or persistent--symptoms, poor control, and secondary consequences). The measures are limited only to reporting consecutively high urine sugar readings because other symptoms/problems are difficult to define clearly for respondents and it is not always true that they should be reported to the physician as soon as they occur. The consecutively high sugar readings are something all consulting physicians agreed should be reported immediately so this operational definition of the Delphi concept has been retained.

REPORTING TO PHYSICIAN ITEMS

DELPHI 59

BE7 The last time you lost consciousness (passed out) from an insulin reaction, did you contact a doctor about it?

Yes, right away
Yes, within a few days
Yes, a week or more later
No
I have never passed out from an insulin reaction

DELPHI 53

BD 13 The last time that you had several urine tests in a row with high sugar readings, did you tell your doctor's office or clinic?

Yes
No
I haven't had several urine tests in a row with high sugar readings

BD13A IF YES; How soon did you tell your doctor's office or clinic?

Right away
Within a few days
A week or more later

BD14 The last time you had one urine test with a 3+ or higher sugar reading, did you tell your doctor's office or clinic?

Yes
No
I haven't had a urine test with a 3+ or higher sugar reading

BD14A IF YES: How soon did you tell your doctor's office or clinic?

Right away
Within a few days
A week or more later

DELPHI 60

BE7 The last time you lost consciousness (passed out) from an insulin reaction, did you contact a doctor about it?

Yes, right away
Yes, within a few days
Yes, within a week or more later
No
I have never passed out from an insulin reaction

SAFMD - DELPHI 59 - REPORTS SEVERE REACTION TO M.D. IMMEDIATELY

RETAINED ITEMS: BE7

DELETED ITEMS: None

ANALYSIS GROUP: Tests urine for sugar

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	No Item	?
Alpha	No Item	
INVALIDITY		
Compliance	.07	.01
VALIDITY		
Informant	.38	.38

PRELIMINARY RECOMMENDATION: It can be assumed that the item has "typical" stability characteristics. The item can be used in combination with one or two newly developed items to measure the intended concept.

SAFMD - DELPHI 60 - REPORTS MODERATE OR GREATER LEVELS OF URINARY KETONES TO PHYSICIAN

RETAINED ITEMS: BD15A (Uses BD15)

DELETED ITEMS: (None)

ANALYSIS GROUP: Tests urine for sugar

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	No Item	?
Alpha		
INVALIDITY		
Compliance Bias	.04	.00
VALIDITY		
Informant	.37	.37

PRELIMINARY RECOMMENDATION: On the basis of reasonably good single-item validity and a low amount of bias, it is assumed that the item has satisfactory single-item stability properties. It can be used in combination with other, newly developed, items to measure the intended concept.

SAFEMD - DELPHI 53 - REPORTS SYMPTOMS TO M.D. (HIGH SUGAR TEST RESULTS)

RETAINED ITEMS: BD13A (Uses BD13), BD14A (Uses BD14)
DELETED ITEMS: (None)
ANALYSIS GROUP: Tests urine for sugar.

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	No Items	
Alpha	.77	.77
INVALIDITY		
Compliance Bias	.24	.11
VALIDITY		
Informant	.47	.47

PRELIMINARY RECOMMENDATION: Satisfactory stability and validity. Can be retained with the limitation that compliance bias must be controlled for in any analysis.

Insulin Medication

Four indexes of insulin taking behavior were constructed and evaluated. The index of whether the patient takes insulin at the appropriate times of day has acceptable properties provided the appropriate times of day are defined by the medical record (rather than by the patient's reported version of his prescription). One item on the retest form is used to construct the candidate scale. This item was mistakenly omitted from the first questionnaire. It is not included in the retest scale used to estimate retest reliability. The index of how accurately the patient measured the insulin dose is not acceptable because of low internal consistency. However, the next table presents the demonstration results. The demonstration score is whether the patient draws up exactly the amount of insulin he said he normally takes in the morning.

Patient Question AB14 (7 day)	Demonstration Correct	Demonstration Wrong
Correct	18	3
Wrong	1	0

The table suggests that behavior variation is not great, and that the few disagreements are due mostly to a patient compliance reporting bias. The small compliance bias did not show up in the questionnaire analysis. We do not feel a good questionnaire measure has been developed but one may not be needed if skill levels in the population are as high as they are in the demonstration.

The single-item* index of self-initiated changes in the insulin regimen has uncertain measurement properties. The major problem experienced was an inability to write items that both capture the intent of the two Delphi concepts and also contain answers that can be scored clearly correct or clearly wrong. Unfortunately, prescriptions for self-initiated changes in the insulin regimen (type, dose, timing) vary a great deal and are not reflected very exactly in the medical records. Records, for example, do not state what dose changes

*Although only one item is scored, it is necessary to ask three questions to obtain the responses.

the patient is expected to make in response to what kinds of subjective symptoms or what patterns of urine test results. Our items are phrased at a very general level and, consequently, may not achieve acceptable measurement standards if a full evaluation is made in future research.

The general item about insulin medication contains a relatively high amount of response bias and should not be used in subsequent research.

INSULIN ITEMS

DELPHI 99

AA17 (Yesterday, or on the last day that you were not sick), When did you inject your insulin?

Before breakfast
Before supper
Other (Please write in:)

AB13 (During the past 7 days, or during the last 7 days that you were not sick), If you take insulin, how many of your insulin injections did you take when you were supposed to?

All of them
Most of them
Some of them
None of them
I did not take insulin

BB10 (During the past 7 days, or during the last 7 days that you were not sick), When did you usually inject insulin?

Before breakfast
Before supper
Other, please write in _____
I did not inject insulin

DELPHI 05

AA19 (Yesterday or on the last day you were not sick), What was the total number of units of insulin you injected each time?

# of units (USE ONE LINE FOR EACH TIME YOU INJECTED INSULIN)

1. _____
2. _____
3. _____

AB14 (During the past 7 days, or during the last 7 days that you were not sick), How many of your insulin doses did you measure the way you were supposed to?

All of them
Most of them
Some of them
None of them
I did not take insulin

DELPHI 08/11

AJ14 Were you told to change your insulin dosage or schedule sometimes--for instance, if you are sick or have an insulin reaction?

Yes
No

AJ16 During the last 30 days, were there any times when you should have made the changes you described above?

Yes
No

AJ16A IF YES: Did you make the changes you should have?

Yes, every time I should have
Yes, most of the time
Yes, some of the time
No

GENERAL

AE10 Think about how and when you took your insulin during the last 30 days. Have you done what your doctor's office or clinic told you to do?

Yes, exactly
Yes, most of the time
Yes, some of the time
Yes, a little of the time
No, not at all

MEDINS - DELPHI 99 - TAKES INSULIN AT APPROPRIATE TIMES

RETAINED ITEMS: AA17, AB13, BB10
DELETED ITEMS: (None)
ANALYSIS GROUP: Insulin dependent

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.73	.67
Alpha	.60	
INVALIDITY		
Compliance Bias	.02	.00
VALIDITY		
Informant	.41	.41

PRELIMINARY RECOMMENDATION: Score patient responses using medical records information for the prescription. Retain existing items.

MEDINS - DELPHI 05 - ACCURATELY MEASURES INTENDED DOSE OF INSULIN

RETAINED ITEMS: AA19, AB14
DELETED ITEMS: AA18
ANALYSIS GROUP: Insulin Dependent

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	No Item	.19
Alpha	.19	
INVALIDITY		
Compliance Bias	-.03	Some*
VALIDITY		
Informant	Item Deleted (Low Variance)	Some*

*See text for discussion of validity determined from demonstration data.

PRELIMINARY RECOMMENDATION: New development research needed. The evaluations here are incomplete but suggest that the items tried are not a good basis for developing the desired scale.

MEDINS - DELPHI 08/11 - CORRECTLY VARIES INSULIN DOSE

RETAINED ITEM: AJ16A (Uses AJ14 and AJ16)

DELETED ITEMS:

ANALYSIS GROUP: Insulin Dependent

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	No Item	
Alpha	No Item	
INVALIDITY		
Compliance Bias	.04	.00
VALIDITY		
Informant	.28	At least .28

*Possibly an underestimate of validity. The patient-informant disagreement is in the direction opposite from the patient compliance bias hypothesis prediction. Informant answers may not be a valid criterion for evaluating patient responses.

PRELIMINARY RECOMMENDATION: New research is needed. It would develop and evaluate items that are specific to dosage changes in response to subjective symptoms or objective urine test results.

MEDINS - GENERAL ITEM

RETAINED ITEM: AE10
DELETED ITEM: (None)
ANALYSIS GROUP: Insulin Dependent

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest (BE2)	.45	.45
Alpha	No Items	
INVALIDITY		
Compliance	.26	.13
VALIDITY		
Informant	No Item	?

PRELIMINARY RECOMMENDATION: Along with other general items, this one probably should not be used further.

Oral Medication

Due possibly to the small number of patients in our sample who reported taking oral medication, we did not develop good measures of taking oral medication. Of primary concern is the extremely low reliability of the answers. We recommend that our questions not be used in future research and that further measurement work with adequate patient samples be undertaken to develop the needed indexes.

ORAL MEDICATION ITEMS

DELPHI 01

AB15 (During the past 7 days (or during the last 7 days that you were not sick), If you take pills to lower your blood sugar, on how many days did you take fewer pills than you were supposed to?

Every day
Most days
Some days
Never
I do not take pills to lower my blood sugar

AF2 (Last 30 days). At any time in the last 30 days, did you ever forget to take your pills?

Yes
No

AF3 (Last 30 days). At any time in the last 30 days, did you ever take fewer pills than you were told to because your diabetes wasn't bothering you?

Yes
No

DELPHI 03

AF4 (During the last year). At any time in the last year did you ever miss taking any pills because you didn't get your prescription filled?

Yes
No

AF5 When did you get your prescription for your pills refilled the last time?

Before I ran out
Within a day or two after I ran out
A week or more after I ran out

GENERAL

AF6 (During the last 30 days). Think about taking your pills during the last 30 days. Have you done what your doctor's office or clinic told you to do?

Yes, exactly
Yes, most of the time
Yes, some of the time
Yes, a little of the time
No, not at all

MEDOR - DELPHI 01 - TAKES CORRECT PILL DOSE

RETAINED ITEMS: AB15, AF2, AF3
DELETED ITEMS: AA14 (low variation)
ANALYSIS GROUP: Takes pills

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	Item Deleted*	
Alpha	.35	.35
INVALIDITY		
Compliance Bias	-.02	.00
VALIDITY		
Informant	Item Deleted**	Unknown

*Low variance in the retest sample

**Low variance, only one informant mentioned a dosage error.

PRELIMINARY RECOMMENDATION: Drop these items. Begin new development research on adequate patient sample.

MEDOR - DELPHI 03 - FILLS PILL PRESCRIPTION

RETAINED ITEMS: AF4, AF5
DELETED ITEMS: (None)
ANALYSIS GROUP: Takes pills.

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.00	.25
Alpha	.33	
INVALIDITY		
Compliance Bias	.08	.01
VALIDITY		
Informant	.26	.26

PRELIMINARY RECOMMENDATION: These items should be dropped. New research is needed.

MEDOR - GENERAL ITEM

RETAINED ITEM: AF6
DELETED ITEMS: (None)
ANALYSIS GROUP: Takes pills

Measurement Characteristic	Correlation	Tentative Interpreation
STABILITY		
Retest	.50	.50
Alpha	No Items	
INVALIDITY		
Compliance Bias	.18	.06
VALIDITY		
Informant	No Item	?

PRELIMINARY RECOMMENDATION: As other general items, this one appears to contain some compliance bias and probably should not be used further.

Urine Testing

Diabetic patients are thought to vary a great deal in whether or not they regularly test their urine for sugar. Our preliminary analysis suggests we have developed a promising index that captures at least some of this variation but also contains too much compliance reporting bias. The index does not have the high stability characteristics expected when many similar items are combined into a single score. This suggests there may be a good deal of (true) day-to-day variation in compliance.

There is a choice of defining the patient's urine testing regimen requirements using the medical record or the patient's recall of the prescription. We have computed this index both ways and recommend that the medical record information be used.

URINE TESTING ITEMS

DELPHI 98

AA8 (Yesterday, or on the last day that you were not sick), Did you test your urine for sugar?

Yes
No

AA9 (Yesterday, or on the last day that you were not sick), At which of these times did you test your urine?

Just before breakfast
Just after breakfast
Just before lunch
Just after lunch
Just before supper
Just after supper
At bedtime

AB9 (During the past 7 days, or the last 7 days that you were not sick), On how many days did you test your urine for sugar?

Every day
Most days
Some days
None of the days

URTEST - DELPHI 98 - TESTS URINE FOR SUGAR

RETAINED ITEMS: AA8, AA9, AB9
DELETED ITEMS: (None)
ANALYSIS GROUP: Tests urine for sugar

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.72	.79
Alpha	.86	
INVALIDITY		
Compliance Bias	.26	.13
VALIDITY		
Informant	.57	.57

PRELIMINARY RECOMMENDATION: Index can be used with little or no further testing provided separate estimates of compliance bias are available to adjust estimates of the mean.

Urine Testing Technique

Our data suggest it is difficult but not impossible to assess the adequacy of a patient's urine testing technique with a self-administered questionnaire. Preliminary analysis suggests that it may be feasible to assess some kinds of technical behavior but at a cost of including many detailed questions in the questionnaire. Unfortunately, one type of technical behavior does not appear to predict other types; it would be incorrect, therefore, to attempt to measure some underlying single dimensions of technical urine testing behavioral competence.

The three-item index of testing at correct times in relation to meals has low validity, moderate reliability, and some bias. It is possible that adding items may improve the measurement properties but further testing will be required. It is also possible that compliance truly varies over time and that the items are reflecting this variation.

The four-item index of storing testing materials properly does not have good measurement characteristics but probably can be improved by a simple rewording of the questions. Suggested new wordings are given in the table.

Three single-item indexes are evaluated. None are satisfactory by themselves but preliminary indications are that further research may yield reasonably good indexes of correct test timing and whether ketone tests are conducted when high urine sugar readings are obtained. It will be more difficult to construct good measures of test interpretation behavior using self-administered questionnaires.

The validity of the test timing item can be suggested by the demonstration data (Table 10). Two urine sugar test trials were taken during the demonstration, one with a solution containing glucose, another with a distilled water solution, tinted yellow. The questionnaire and two test timing demonstrations yield roughly the same classifications of correct and incorrect timing when both measures are available. The cross-classified results tend to support the finding of moderate validity levels and little or no compliance bias in the patient answers.

The general item about urine testing technique (AD12) should be dropped from further use because of low between-patient variation and because the existing variation was correlated with compliance response bias.

We did not attempt to measure two concepts that the expert panel rated important: uses medicine dropper properly (for Clinitest sugar test) and accurately matches test color to the reference chart. The latter is somewhat covered by our approach to Delphi concept 28A.

Table 10

CROSS-CLASSIFICATION FREQUENCIES OF
QUESTIONNAIRE TEST TIMING RESPONSES
WITH EACH OF TWO DEMONSTRATION TEST RESULTS

AD5 - Wait Correct Seconds

Demonstration		Yes	No	Total
Wait Correct	Yes	4	2	6
Seconds	No	4	7	11
Test 1	Total	8	9	

Demonstration		Yes	No	Total
Wait Correct	Yes	7	1	8
Seconds	No	2	7	9
Test 2	Total	9	8	

URINE TESTING TECHNIQUE ITEMS

DELPHI 26

AA9 (Yesterday, or on the last day that you were not sick), At which of these times did you test your urine?

Just before breakfast
Just after breakfast
Just before lunch
Just after lunch
Just before supper
Just after supper
At bedtime

AB10 (During the past 7 days or last 7 days that you were not sick), When you tested your urine for sugar, did you test it at the time of day you were supposed to?

Yes, always
Yes, usually
Yes, some of the time
No
I did not test my urine

AD3 (The last time you tested your urine), Had you eaten a meal or snack during the hour before you conducted the test?

Yes
No

DELPHI 23

AD10A Where are your urine testing materials now--are they in a cool place?

Yes
No

AD10B Where are your urine testing materials now--are they in a dark place?

Yes
No

AD10C Where are your urine testing materials now--are they in a place where they might get damp?

Yes
No

DELPHI 21

AD4 (The last time you tested your urine), What was the name of the test you used?

Testape
Clinitest
Diastix
Keto-diastix
Clinistix
Other (Write in Name)

AD5 (The last time you tested your urine), How long did you wait before you compared the color of the test to the color chart?

Write in number of seconds _____
I don't know

DELPHI 28A - Also uses AD4

AD6 (The last time you tested your urine), What color was the test reaction?

Write in color
I don't know

AD7 (The last time you tested your urine), What was the sugar reading for that test?

Write in result
I don't know

DELPHI 28B - Also uses AD4

AD11 (In the last 30 days). Please list below the three highest sugar readings you got in the last 30 days.

Sugar Readings
1. _____
2. _____
3. _____

Then, please list the acetone test results for these readings.

Acetone Readings	(IF YOU DID NOT TEST FOR ACETONE, CHECK BELOW)
1. _____	___
2. _____	___
3. _____	___

GENERAL ITEM

AD12 (In the past 30 days). Now, think about how and when you tested your urine during the past 30 days. Have you done what your doctor's office or clinic told you to do?

Yes, exactly
Yes, most of the time
Yes, some of the time
Yes, a little of the time
No, not at all

URTEC - DELPHI 26 - TESTS URINE AT CORRECT TIMES IN RELATION TO MEALS

RETAINED ITEMS: AB10, AD3, AA9
DELETED ITEMS: AD2 (Could not be scored meaningfully)
ANALYSIS GROUP: Tests urine for sugar

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.38	.41
Alpha	.44	
INVALIDITY		
Compliance Bias	.18	.06
VALIDITY		
Informant	.22	.22
	.07	

PRELIMINARY RECOMMENDATION: Need more items to tap this part of technical urine testing behavior reliably or else determine that the low answer "stability" reflects true day-to-day variation. Users may need measures of compliance bias for interpretation of mean scores.

URTEC - DELPHI 23 - STORES URINE TEST MATERIALS SO POTENCY IS MAINTAINED

RETAINED ITEMS: AD10A, AD10B, AD10C

DELETED ITEMS: AD10D (low variation between patients)

ANALYSIS GROUP: Tests urine for sugar.

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	No Items	
Alpha	.33	.33
INVALIDITY		
Compliance Bias	.08	.01
VALIDITY		
Informant	No Items	Low

PRELIMINARY CONCLUSIONS: The items appear to need revision and new testing. Recommended revised items:

10. Where are your urine testing materials now ?

A. Where they are warmed by the sun or a heater... Yes..... 1
No..... 2

B. In a dark place..... Yes..... 1
No..... 2

C. Where they are kept dry..... Yes..... 1
No..... 2

D. Where the chemicals can get lots of fresh air.. Yes..... 1
No..... 2

URTEC - DELPHI 21 - WAITS CORRECT TIME BEFORE READING RESULT

RETAINED ITEMS: AD5 (uses AD4)
DELETED ITEMS: (None)
ANALYSIS GROUP: Tests urine for sugar.

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.69	.69
Alpha		
INVALIDITY		
Compliance Bias	-.12	.03
VALIDITY		
Informant	No Items	Probably Acceptable*

*See text for discussion of validity determined from demonstrations.

PRELIMINARY RECOMMENDATION: Item can be retained. Answers are reasonably stable, minimally biased, and the small sample demonstration study suggests an acceptable amount of criterion validity. The validity conclusion needs confirmation in future research.

URTEC - DELPHI 28A - INTERPRETS URINE TESTS CORRECTLY

RETAINED ITEMS: AD7 (Uses AD4 and AD6)
DELETED ITEMS: (None)
ANALYSIS GROUP: Tests urine for sugar

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.38	.38
Alpha		
INVALIDITY		
Compliance Bias	.05	.00
VALIDITY		
Informant	No Items	?

PRELIMINARY RECOMMENDATION: A difficult behavior category to try to measure in questionnaire research. Further development and testing is needed but may not result in finding appropriate measures.

URTEC - DELPHI 28B - TESTS FOR KETONES IF SUGAR TEST RESULT HIGH

RETAINED ITEMS: AD11 (Uses AD4)
DELETED ITEMS: AD8 (Uses AD4, 6, 7)
ANALYSIS GROUP: Tests urine for sugar.

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.34	.34
Alpha		
INVALIDITY		
Compliance Bias	-.11	.02
VALIDITY		
Informant	No Items	?

PRELIMINARY RECOMMENDATION: Needs further evaluation, sample sizes are too small for adequate evaluation in this research.

URTEC - GENERAL ITEM

RETAINED ITEM: AD12
DELETED ITEMS: (None)
ANALYSIS GROUP: Tests urine for sugar

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.40	.40
Alpha	No Items	
INVALIDITY		
Compliance Bias	.25	.12
VALIDITY		
Informant	No Item	?

PRELIMINARY RECOMMENDATION: The item should be dropped.

Recording Urine Test Results (URREC)

The index of keeping urine test records is based on 3 items asked for the first time on the "retest" questionnaire. It shows some acceptable characteristics but must be scored using medical record information about the recordkeeping requirements. It is also affected importantly by compliance reporting bias.

The index does not attempt to get at the complex recordkeeping behavior implicit in the Delphi item. We did not attempt to measure whether the patient records the test result accurately nor did we try to define what is an adequate sample of test results. The index measures whether all or some results are recorded (depending on the prescription). The panel experts were unable to define what was an adequate sample of results and we chose not to ask physicians to define this for each sampled patient.

URINE TEST RECORDING ITEMS

DELPHI 27

BA8 (Yesterday, or on the last day that you were not sick), Did you write down your urine test results?

Yes, all of them
Yes, some of them
No
I did not test my urine

BB7 (During the past 7 days, or during the last 7 days that you were not sick), How many of your urine test results did you write down?

All of them
Most of them
Some of them
None of them
I did not test my urine

BD8 (The last time you tested your urine), Did you write down the test result?

Yes
No

URREC - MODIFIED DELPHI 27 - RECORDS URINE TEST IF PRESCRIBED

RETAINED ITEMS: BA8, BB7, BD8 (All scoredusing D20)

DELETED ITEMS: (None)

ANALYSIS GROUP: MD prescribes recording test results

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	No Items	.76
Alpha	.76	.76
INVALIDITY		
Compliance Bias	.21	.08
VALIDITY		
Informant (Delphi)	.48	.48

PRELIMINARY RECOMMENDATION: Retain. Satisfactory stability and validity but will need independent measure of compliance bias to interpret means.

Eats Prescribed Meals at Prescribed Times (NUTEAT)

The first nutrition measures discussed are about general meal timing and content behavior. Index measures of the separate parts of the concept are presented to facilitate discussion; however, the parts may be combined into a single index that is scored according to patient reports of prescribed eating requirements.

The "eats prescribed meal" measure and both versions of the timing measure contain moderate amounts of compliance reporting bias. Scoring behavior using the meal timing requirements in the medical record does not reduce the response bias variance. If the separate indexes are combined, the resulting measure will have a substantially higher proportion of response bias variance than the component indexes individually (but also a higher proportion of true score). Neither index has an appropriately high informant correlation. The general item index of nutrition behavior also contains substantial amounts of compliance response bias.

New measurement approaches might be tried instead of further work aimed at developing these scales.

GENERAL MEAL CONTENT AND TIMING ITEMS

DELPHI 33A

AA3 (Yesterday, or on the last day that you were not sick), Did you eat or drink something that you were not supposed to for any meal or snack?

Yes
No

AB3 (During the past 7 days, or during the last 7 days that you were not sick), On how many days did you eat or drink something you were not supposed to?

Every day
Most days
Some days
Never

AB4 (During the past 7 days, or during the last 7 days that you were not sick), In general, during these 7 days, would you say that you ate more than you should have, less than you should have, or about the right amount?

I ate more than I should have
I ate less than I should have
I ate about the right amount

DELPHI 33B

AA1 (Yesterday, or on the last day that you were not sick), Did you skip any meals?

Yes
No

AA3 (Yesterday, or on the last day that you were not sick), At which of these times did you eat a snack?

Between breakfast and lunch
Between lunch and supper
After supper or at bedtime
I did not eat any snacks

AB1 (During the past 7 days, or during the last 7 days that you were not sick) On how many days did you skip eating a meal or a snack?

Every day
Most days
Some days
Never

AB2 (During the past 7 days, or during the last 7 days that you were not sick), Did you eat any meals or snacks later than you should have?

Yes
No

GENERAL ITEM

AG4 (Past 30 days). Think about when and what you have eaten during the past 30 days. Have you done what your doctor's office or clinic told you to do?

Yes, exactly
Yes, most of the time
Yes, some of the time
Yes, a little of the time
No, not at all

NUTEAT - DELPHI 33A - EATS PRESCRIBED MEALS

RETAINED ITEMS: AA3, AB3, AB4
DELETED ITEMS: (None)
ANALYSIS GROUP: All patients who answered.

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.61	
Alpha	.61	.61
INVALIDITY		
Compliance Bias	.32	.19
VALIDITY		
Informant (Delphi)	.29	.29

PRELIMINARY RECOMMENDATION: Marginal index because of large amount of compliance bias and low validity. Needs further development and testing or should be abandoned in favor of another measurement approach.

NUTEAT - DELPHI 33B - EATS AT PRESCRIBED TIMES

RETAINED ITEMS: AA1, AA2, AB1, AB2

DELETED ITEMS: (None)

ANALYSIS GROUP: All patients who answered.

	PATIENT VERSION		RECORD VERSION	
STABILITY	r	Tentative Interpretation	r	Tentative Interpretation
Retest	.62		.48	
Alpha	.46	.54	.34	.41
INVALIDITY				
Compliance Bias	.21	.08	.22	.09
VALIDITY				
Informant (Delphi)	.24	.24	.25	.25

PRELIMINARY INTERPRETATION: Both versions appear heavily influenced by compliance desirability reporting bias. Neither version achieves a satisfactory informant correlation. Further development and evaluation work is needed to construct better indexes of the concept.

NUTEAT - GENERAL ITEM

RETAINED ITEM: AG4
DELETED ITEMS: (None)
ANALYSIS GROUP: All who responded

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.60	.60
Alpha	No Items	
INVALIDITY		
Compliance Bias	.34	.22
VALIDITY		
Informant	No Item	?

PRELIMINARY RECOMMENDATION: Drop item from further use due to high compliance bias variance.

Technical Food Preparation Behavior

The attempts made to assess the technical aspects of food preparation behavior were only marginally successful. New methods of measurement probably should be tried. The index of weighing or measuring food portions has good stability. Its correlation with the informant information is low when all informant responses are included. When the analysis is limited to about 200 cases where the informant observes patient meal behavior at least daily, the size of the validity correlation increases substantially although it does not get as high as desirable. The index of whether the patient's portion sizes are estimated also has reasonable stability properties but lower validity. The lower validity is not apparently due to lack of informant knowledge.

The index of whether a food exchange list is used to prepare meals is somewhat troublesome. First, the choice of the technical term "food exchange list" was unfortunate. Patients apparently took this to mean almost any set of rules for preparing meals and not the highly detailed manuals that are based on recommendations of the American Diabetes Association. (Two of the four facilities providing the patient sample do not use the food exchange list method of meal planning; yet many of these patients claimed to be using the list.) The index variation contains a moderate amount of compliance bias. Also the informant correlation is not high. Valid measures of the construct might be developed if the questions contain better definitions of what an exchange list is.

The index of accounting for alcohol in the meal plan has good measurement characteristics. Despite small patient sample sizes, the index has very good stability, only very small amounts of compliance bias and reasonable validity levels. The informant validity correlation is substantially higher for the retest item than for the main questionnaire item but the reason for the discrepancy is not clear.

FOOD PLANNING ITEMS

DELPHI 44

AA4 (Yesterday, or on the last day that you were not sick), Did you measure or weigh the food that you ate?

Yes, all or most of if
Yes, some of it
No, none of it

AB6 (During the past 7 days, or during the last 7 days that you were not sick), For how many meals or snacks did you measure or weigh the food that you ate?

For all of them
For most of them
For some of them
For none of them

DELPHI 44E

AA6 (Yesterday, or on the last day that you were not sick), If you did not weigh or measure your food, did you try to estimate the size of each serving?

Yes
No

AB7 (During the past 7 days, or during the last 7 days that you were not sick), When you did not weigh or measure your food, how often did you estimate the size of the servings so that you would not eat too much or too little?

For all of my meals and snacks
For most of my meals and snacks
For some of my meals and snacks
For none of my meals or snacks

DELPHI 41

AB5 (During the past 7 days, or during the last 7 days that you were not sick), How many meals or snacks were prepared according to a food exchange list?

All of them
Most of them
Some of them
None of them

AG2 (In the past 7 days). During the past 7 days, did you use a food exchange list to plan how much you could eat of each of the following foods?

The fruit that you ate?
Yes, always; yes, sometimes; no; I did not eat or drink this.
The meat that you ate?
Yes, always; yes, sometimes; no; I did not eat or drink this.
The desserts that you ate?
Yes, always; yes, sometimes; no; I did not eat or drink this.
The bread that you ate?
Yes, always; yes, sometimes; no; I did not eat or drink this.
The lettuce or raw tomatoes that you ate?
Yes, always; yes, sometimes; no; I did not eat or drink this.
The potato chips or other snacks like that which you ate?
Yes, always; yes, sometimes; no; I did not eat or drink this.
The alcoholic beverages that you drank?
Yes, always; yes, sometimes; no; I did not eat or drink this

DELPHI 42

AA7 (Yesterday, or on the last day that you were not sick), If you drank any alcoholic beverages, did you count them in your food plan?

Yes
No
I did not drink any alcoholic beverages

AB8 (During the past 7 days, or during the last 7 days that you were not sick), If you drank any alcoholic beverages during these 7 days, how often did you count them in your food plan?

Every time
Most of the time
Some of the time
None of the time
I did not drink any alcoholic beverages

DELPHI 44W - WEIGHS OR MEASURES PORTIONS

RETAINED ITEMS: AA4, AB6
DELETED ITEMS: (None)
ANALYSIS GROUP: All who answered

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.64	.73
Alpha	.82	
INVALIDITY		
Compliance Bias	.17	.06
VALIDITY		
Informant	.28-.44*	.44

*When limited to approximately 200 informants who observe the patient's meal behavior daily, the correlation is .44.

PRELIMINARY RECOMMENDATION: The index has marginal properties and probably needs to be redeveloped. It contains some bias and its validity is uncertain.

NUTTEC - DELPHI 44E- ESTIMATES PORTIONS

RETAINED ITEMS: AA6, AB7
DELETED ITEMS: (None)
ANALYSIS GROUP: All who answered

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.57	.62
Alpha	.67	
INVALIDITY		
Compliance Bias	.10	.02
VALIDITY		
Informant	.21-.28*	.28

*The .28 correlation is obtained using only informants who observe the patient's meal behavior at least daily.

PRELIMINARY RECOMMENDATION: The index has marginal validity properties. New research is recommended.

NUTTEC - DELPHI 41 - HAS AND USES FOOD EXCHANGE LIST

RETAINED ITEMS: AB5, AG2 (subitems a, b, c, d, f, g)*
DELETED ITEMS: (None)
ANALYSIS GROUP: All who answered

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.69	.66
Alpha	.62	
INVALIDITY		
Compliance Bias	.25	.12
VALIDITY		
Informant	.28	.28

*Subitem e is a compliance bias validation item and is excluded from the index.

PRELIMINARY RECOMMENDATION: Will require a good deal of further research to develop acceptable measures of this concept.

NUTTEC - DELPHI 42 - IF CONSUMES ALCOHOL, INCLUDES IN PLANNING CALCULATIONS

RETAINED ITEMS: AA7, AB8

DELETED ITEMS: (None)

ANALYSIS GROUP: All who answered

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	.60	.72
Alpha	.83	
INVALIDITY		
Compliance Bias	.10	.02
VALIDITY		
Informant	.24*	At least .24

*The correlation between the retest item and the informant item is .60.

PRELIMINARY RECOMMENDATION: Retain the index.

Weight Control

The last nutrition index measures weight control based on self-reported weight at the present time and at two times in the past. Although self-reported weight in general population studies is sometimes biased toward a perceived ideal weight, no evidence for this kind of bias is found in our diabetic patient group. The index has good stability and validity properties and appears to contain almost no compliance reporting bias.

WEIGHT CONTROL ITEMS

DELPHI 37

A7 About how much do you weigh now?

_____ pounds

A8 About how much did you weigh 6 month ago?

_____ pounds

A9 About how much did you weigh one year ago?

_____ pounds

BG5 According to your doctor, do you weigh too much, too little, or about the right amount?

Too much
Too little
About the right amount
I don't know

NUTWGT - DELPHI 37 - ACHIEVES AND MAINTAINS IDEAL BODY WEIGHT (± 10%)

RETAINED ITEMS: BG5; A7, A8, A9 Scored using AJ7,
DELETED ITEMS: (None)
ANALYSIS GROUP: All who answered

Measurement Characteristic	Correlation	Tentative Interpreation
STABILITY		
Retest	See Below*	
Alpha	.77	.77
INVALIDITY		
Compliance Bias	.10	.02
VALIDITY		
Informant (Delphi)	.73	.73

*Weight history was not asked at retest although current weight was. The retest correlation of the variable, "current weight is within ± 10 percent of perceived ideal," was .90.

PRELIMINARY RECOMMENDATION: Use existing items because of satisfactory stability, apparent high validity, and low amount of bias.

Foot Care

The foot care index includes questions about whether the adult, insulin dependent patient periodically inspects his feet for possible problems. The index has good internal consistency characteristics but contains some bias and the bias attenuates the validity coefficient. Additional research will be needed to develop an index which is less subject to bias. Our questions did not attempt to measure toenail cutting behavior nor assess how well the patient's shoes fit.

FOOT CARE ITEMS

DELPHI 54

BA11 (Yesterday), Did you check your feet for any signs of problems?

Yes
No
I do not need to check my feet for problems

BB11 (7 days), How often did you check your feet for signs of problems?

Every day
Most days
Some days
None of the days

FEET - DELPHI 54 - INSPECTS FEET DAILY, CUTS NAILS CORRECTLY, AND WEARS SHOES THAT FIT WELL

RETAINED ITEMS: BA11, BB11

DELETED ITEMS: (None)

ANALYSIS GROUP: Insulin dependent, diagnosed diabetic 10 or more years ago, and is over 12 years old.

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	No Items	
Alpha	.80	.80
INVALIDITY		
Compliance Bias	.20	.08
VALIDITY		
Informant	.30	.30

PRELIMINARY RECOMMENDATION: Needs more developmental research to better understand the magnitude of response bias and the low to moderate informant correlation.

Obtaining Medical Care (CAREMD)

An index of whether the patient informed all of his medical care providers that he has diabetes was developed. The evaluations that were done (internal consistency reliability and correlation with compliance response bias) indicate good measurement properties.

The Delphi panel also recommended that we develop measures of whether the patient visits medical care providers as frequently as each provider recommends. Questionnaire space limitations prevented us from undertaking the large task of developing accurate measures of how often specialist visits are prescribed for each patient and, then, trying to measure (independently) how often the recommendations were followed for each type of provider.

CAREMD - DELPHI 52 - INFORMS ALL MEDICAL CARE PROVIDERS OF DIABETIC CONDITION

RETAINED ITEMS: BH6 (5 parts)

DELETED ITEMS: (None)

ANALYSIS GROUP: All who answered

Measurement Characteristic	Correlation	Tentative Interpretation
STABILITY		
Retest	No Items	
Alpha	.79	.79
INVALIDITY		
Compliance Bias	-.03	.00
VALIDITY		
Informant (Delphi)	No Items	

PRELIMINARY RECOMMENDATION: The measure has satisfactory stability and bias characteristics and should be retained pending further evaluation of its validity characteristics in future research.

III. KNOWLEDGE

SUMMARY OF FINDINGS

Twenty-one multi-item knowledge subtests hypothesized to measure important knowledge concepts (Delphi item groupings) were evaluated using four criteria derived from test theory. Twelve knowledge subtests containing from two to ten items each satisfied all four criteria. The successful tests cover parts of each of the major self-care categories: safety, medication, urine testing, and nutrition.

The first criterion (item convergent validity), which required that items in the same subtest correlate positively with the subtest total score, was met by 14 of the 21 hypothesized knowledge subtests. All but one of these subtests satisfied the second criterion, which was that the subtest exhibit satisfactory internal consistency. The third and most stringent criterion (item discriminant validity) required that items correlate higher with their hypothesized subtest than with other knowledge subtests. This criterion was satisfied by 13 of the 21 subtests. Revisions in two subtests were necessary to satisfy the third criterion. None of the 13 subtests correlated significantly with self-reports of tendencies to guess rather than say "don't know" (the fourth criterion).

Four of the hypothesized subtests were improved by eliminating items that accounted for a disproportionately large number of item discriminant validity errors. Elimination of one or two items from these four subtests substantially improved score reliability and made it possible for retained items in two subtests to satisfy other test construction criterion (e.g., item discriminant validity).

The percentage of correct responses per item tended to be higher than optimal for purposes of discriminating among individuals throughout the range of possible scores. If this trend is replicated in other populations, there may be value in revising some items or item scoring rules to increase difficulty and to obtain more information about knowledge throughout the range of possible knowledge scores. This strategy would particularly improve discriminations among more knowledgeable inviduals.

Studies of correlations among knowledge subtests revealed a general knowledge factor and four interpretable knowledge dimensions. Tests constructed to measure these dimensions performed very well in relation to scaling criteria and their scores achieved reliability coefficients ranging from 0.65 to 0.90. These four knowledge tests may be preferred by some researchers.

Ten knowledge concepts identified as important by the Delphi panel could not be represented by multi-item knowledge subtests based on current study items. Two of these areas were measured by only one item each; tests to measure them should be lengthened. Many other concepts were represented by only two items that were not internally consistent; tests of these concepts should be developed and evaluated using more items. Some areas represented by more items were not internally consistent as hypothesized; these should be evaluated further with different strategies for grouping items and possibly with additional items. Lack of response variability prevented evaluations of some knowledge tests; these should be revised to achieve greater score variability. Limited validity analyses of a few scales were carried out using in-clinic demonstration data. Neither internal nor predictive validity of any of the tested scales was high.

METHODOLOGICAL NOTES

Scaling

Multi-item knowledge tests and attitudinal scales, if properly constructed, serve very useful purposes. First, they reduce the number of variables that must be scored and interpreted for purposes of hypothesis testing (i.e., the number of variables is reduced from the number of items to the number of multi-item tests or scales). The tradeoff between few and many measures is one of simplicity versus loss of information. If too many measures are constructed, simplicity is lost; the gain is more focused measurement. If too few measures are constructed, information about specific constructs may be lost and the analysis may prove difficult due to difficulties in interpreting scores.

Multi-item measures also increase reliability by pooling variance that is common across items. The working hypothesis was that knowledge items in the same Delphi grouping would contain some common variance. Thus, combining these items increases reliability of measurement by replicating (across items) measurement of the common variance. Further, because each item represents the construct defined by the grouping, it is not necessary to measure all possible items in the grouping to estimate knowledge of the construct; a good sample of items will suffice. It was also hypothesized that items in each grouping would contain some unique reliable variance (i.e., variance relevant to the construct but not shared with any other item). By combining items that share common variance and also contribute unique (true construct) variance, validity of measurement may be improved. In other words, measurement of the construct is more comprehensive because more aspects of the construct are represented by the items. Thus, properly constructed multi-item measures should also have greater predictive validity.

The scaling of knowledge items (i.e., the construction of multi-item knowledge subtests) was carried out in three steps. First, items were grouped according to behavioral content areas (Delphi grouping) and more global concepts to define hypothesized subtests (see Table 11 for a summary. It contains 23 groupings, two of which contain only one item and are not evaluated further.)

The second step tested item convergent and discriminant validity. The correlation between a knowledge item and its subtest score was examined to evaluate how well the item measured the construct defined by the subtest (item convergent validity). Because some knowledge subtests were very short (i.e., contained few items) these item-test correlations were corrected for overlap (i.e., they were corrected for bias due to inclusion of the relevant item in the subtest total). Thus, each convergent validity correlation was actually a correlation between an item and the sum of other items in the same subtest.

Next, the criterion of discriminant validity was tested by examining how often each item correlated more highly with its own scale score (convergent correlation) than with other scale scores

(discriminant correlations). A "probable" scaling error occurred when a discriminant correlation is within one standard error of the convergent correlation. A "definite" scaling error is observed when a discriminant correlation is more than one standard error higher than the convergent correlation. A "success" occurs when the convergent correlation is more than one standard error higher than the discriminant correlation.

Third, product-moment correlations among items were factor analyzed; specifically, principal components were extracted and rotated using Kaiser Varimax criteria. The factor analysis served two purposes. It provided another empirical test of hypothesized item groupings (i.e., another test of whether items in the same hypothesized grouping were more highly interrelated with each other than with items in other groupings). The factor analysis also made it possible to test for item groupings not hypothesized.

Distributions of responses were examined to determine whether variability was sufficient for purposes of testing scaling assumptions. For the dichotomous knowledge scores (correct-incorrect), we examined the percentage correct. The latter statistic is usually referred to as the item "difficulty level." Whereas some easy and some difficult items are desirable, norm-referenced knowledge tests usually prove most reliable and valid when difficulty levels are in the 40 to 60 percent range.

Reliability and Validity

Following construction of multi-item subtests to measure knowledge, internal-consistency and test-retest reliability coefficients were computed along with descriptive statistics (i.e., means and standard deviations) for subtest scores. The former were evaluated to determine whether multi-item subtests improved reliability over what would be expected with single-item measures; the former and the latter were examined to determine whether subtests would facilitate group comparisons as desired.

Table 11

SUMMARY OF FORM A AND B KNOWLEDGE ITEMS BY GLOBAL CONCEPTS AND 23 DELPHI-ITEM GROUPINGS

Global Concepts and Delphi-Item Groupings	Number of Items Form A	Form B
Safety Self-Care (SAFSLF)		
11K Adjusts Insulin Dosage[a]	1	0
34K Changes Eating Time	5	1
36K Eats for Hypoglycemia	11	6
58K Carries Emergency Glucose	2	1
61K Carries Diabetic ID	3	2
62K Tells Others about Hypoglycemia	2	1
Safety MD Care (SAFMD)		
53K Reports Symptoms to MD	0	3
59K Reports Insulin Reactions to MD	0	1
Oral Medications (MEDOR)		
1K Pill Dosage	5	1
3K Pill Prescriptions	2	1
Insulin Taking (MEDINS)		
5K Measures Insulin Correctly	4	3
99K Insulin at Correct Times	4	2
Urine Testing--Technical Aspects (URTEC)		
21K Times Test Correctly	4	4
23K Stores Test Materials Properly	2	1
26K Tests Urine Before Meals	2	2
28K Interprets Test Correctly	9	8
Urine Testing--Record Keeping (URREC)		
27K Keeps Adequate Record	0	2
Nutrition--Eating (NUTEAT)		
33K Prescribed Meals at Prescribed Times	7	2
Nutrition--Technical (NUTTEC)		
41K Uses Food Exchange List	9	2
42K Plans Alcohol Intake	2	1
44K Weighs/Measures Food	4	2
Nutrition--Body Weight (NUTWGT)		
37K Achieves/Maintains Correct Weight	2	0
Foot Care (FEET)		
54K Inspects Feet Daily	0	5
TOTALS	80	51

[a]Numbers refer to the concepts rated important by the Delphi panel (see Appendix A to Summary Report). "K" indicates that knowledge items are used.

Correlations among subtest scores were also factor analyzed to identify the best measures of more global knowledge concepts and to determine whether higher order knowledge factors correspond to global knowledge concepts. Other hypotheses relevant to validity and sources of response bias were also tested. For example, correlations between knowledge subtests and indirect measures of response bias were examined. Included among the latter were measures of favorable (compliance) reporting bias (based on a ten-item scale constructed from Section C of Form A), and a single-item measure of tendencies to guess as opposed to responding "don't know" (AK3) when answers to questions are not known. Construct validity was tested by correlating knowledge subtests with respondent educational level to test the hypothesis that scores for valid measures of knowledge should increase with education. Finally, knowledge scores were compared to patient skill demonstrations from which further impressions of internal and predictive validity are obtained.

Sample Selection for Item Scaling Studies

The use of questionnaire skip patterns produced very large differences in sample sizes across knowledge items. Therefore, it was not desirable to perform scaling analyses using all available data because differences between inter-item correlations could result both from sample differences and differences in the strength of relationships between constructs measured by the items. For purposes of making scaling decisions we were, of course, most interested in the latter.

Patterns of missing responses (due to skip patterns) across items were examined to determine an appropriate analysis sample for purposes of testing scaling assumptions for hypothesized attitude and knowledge item groupings. Two strategies (i.e., elimination of items and respondents) were used to arrive at the largest number of items completed by the largest number of respondents. The largest sample of persons who completed the same items was composed of persons who reported taking insulin, testing urine, and using a meal plan. With the exception of missing responses that did not result from skip patterns, complete data for nearly all knowledge items were available for these respondents. However, a disproportionately large number of

missing responses (not due to skip patterns) was observed for some respondents in this homogeneous sample. Those respondents who left more than one-fourth of the attitude or knowledge items blank (11 and 18 items or more for attitude and knowledge batteries, respectively) were also eliminated before the scaling analyses. For most analyses, this second sample reduction lowered sample sizes to 209 respondents for knowledge items. Missing responses for persons retained were estimated before testing scaling assumptions.* As described below, most of these respondents were returned to the analysis sample after scaling decisions had been completed (e.g., they were included when reliability was estimated for four general knowledge tests).

In addition to eliminating respondents, as described above, some items were eliminated for either of the following reasons: relatively few responses due to skip patterns, or lack of response variance.

Following completion of the scaling evaluations of hypothesized knowledge subtests (the first round of analyses) and after ad hoc revisions suggested by those evaluations, all available data were used to estimate descriptive statistics and reliability coefficients for knowledge subtests. At this stage in the analysis, persons having more than one-fourth missing data were returned to the data file and missing responses were estimated using the procedure previously defined (i.e., the person mean for completed items in the same subtest served as the estimate of the missing item score). This procedure made it possible to increase the sample size from 209 to 310 for studies of the four global knowledge subtests. However, because persons who failed to answer all items in any one subtest were also excluded at this stage, it was not possible to increase the same size for purposes of estimating descriptive statistics and reliability coefficients for the 21 knowledge subtests.

All associations between items were estimated using the product-moment formula. Cross-tabulations were inspected to insure that this

---

*Estimates for missing items were computed from responses to other (completed) items in the same scale. Specifically, the respondent mean score for completed items in the same scale served as the score estimate for the missing item. If all items in a given scale were missing, the respondent was eliminated. Very few (less than 10) respondents were eliminated for the latter reason.

statistic well represented the associations involved. With few exceptions, which will be noted, this appeared to be the case.

Scoring Knowledge Items

All knowledge items were scored dichotomously (one if correct, zero otherwise) before scaling studies and before estimating reliability for item and subtest scores. Table III-A-1 in the annex to this section contains a summary of information about 107 knowledge items not requiring record checks. Variable labels for these items, definitions, and correct answers are presented. Table III-A-2 in the annex presents labels and definitions for 23 knowledge items that required record checks to determine correct answers.

RESULTS

Item Difficulty Levels

Before scaling knowledge items, the difficulty level (i.e., the percentage of correct responses) was computed for each item using data for the total sample and the sample used in the first round of measurement evaluations. Overall, approximately 68 percent of the responses were correct; about 14 percent were left unanswered (and were counted incorrect), and about 18 percent were answered incorrectly (see Table 12). Difficulty levels for 129 knowledge items in the total sample ranged from 14 to 98 percent and the median was 79. From a psychometric point of view, difficulty levels in the 0.40 to 0.60 range are desirable for the great majority of items. The majority of difficulty levels did not fall in the ideal range and there is considerable room for improvement of these knowledge measures by revising scoring rules or items to yield more nearly symmetrical item score distributions. Such revisions would also make it possible to evaluate some items that were eliminated from the scaling analyses due to lack of score variability.

Reliability of Item Scores

Test-retest reliability for single-item scores could be estimated for 21 knowledge items repeated (without substantial variation in item wording) in Forms A and B. These items represented 12 different Delphi item groupings. Test-retest product-moment correlations are presented in Table 13. Test-retest coefficients for these items ranged from a high of 0.76 to a low of -0.33; the median of 21 coefficients was 0.45.

Table 12

FREQUENCY AND PERCENTAGE OF CORRECT RESPONSES FOR KNOWLEDGE ITEMS, TOTAL SAMPLE AND SAMPLE USED IN SCALING ANALYSES

Delphi Groupings/Items	Total Sample f	%	N	Analysis Sample f	%	N
34K[a] Changes Eating Time						
AH14ZA[b]	292	74.3	393	179	89.9	199
AH14ZB	334	78.2	427	179	86.9	206
AH14ZC	298	77.2	386	171	85.9	199
AH14ZD	261	65.4	399	155	77.1	201
BI12	261	70.5	370	148	86.5	171
RKANC34A[c]	87	33.9	257	64	43.0	149
36K Eats for Hypoglycemia						
AH12ZA	177	42.0	421	104	52.0	200
AH12ZB	413	93.0	444	203	98.5	206
AH12ZC	314	74.8	420	160	78.4	204
AH12ZD	184	45.4	405	102	51.3	199
AH12ZE	282	67.0	421	142	68.9	206
AH12ZF	352	80.7	436	178	86.0	207
AH12ZG	213	49.3	432	134	65.4	206
AH13ZA	305	82.0	372	183	92.0	199
AH13ZB	289	78.5	368	176	88.4	199
AH13ZC	436	95.6	456	205	98.1	209
AH13ZD	288	78.0	369	172	87.8	196
BI10	348	93.0	374	170	98.3	173
BI11Z1R	329	83.3	395	162	91.5	177
BI11Z2R	329	83.3	395	169	95.5	177
BI11Z3R	310	78.5	395	144	81.4	177
BI11Z4R	259	65.6	395	135	76.3	177
BI11Z5R	298	75.4	395	152	85.9	177
58K Carries Emergency Glucose						
AA13	349	74.6	468	182	90.1	192
AH1	444	95.9	463	206	98.6	209
BI2	316	82.5	383	161	91.5	176
61K Carries Diabetic ID						
AA11	387	82.9	467	188	91.7	205
AH4	410	88.6	463	197	94.7	208
AH10	443	96.7	458	205	98.6	208
BI4	373	97.9	381	174	99.4	175
BI9	372	97.9	380	173	99.4	174
62K Tells Others about Hypoglycemia						
AH3	321	70.7	454	155	74.9	207
AH9	401	90.3	444	198	97.1	204
BI8	328	91.6	358	167	97.7	171

Table 12 continued

Delphi Groupings/Items	Total Sample f	%	N	Analysis Sample f	%	N
11K Adjusts Insulin Dosage						
DKANC11A	163	58.8	277	114	63.3	180
53K Reports Symptoms to MD						
BI17	309	79.6	388	139	78.5	177
BI18	226	58.2	388	125	70.6	177
BI19	370	94.4	392	171	96.6	177
59K Reports Insulin Reactions to MD						
BI16	369	96.9	381	172	97.7	176
1K Knows Pill Dosage						
AH16	251	77.2	325	102	73.9	138
AH17	246	76.9	320	93	67.9	137
BI13	180	56.6	318	59	43.4	136
RKANDØ1B	37	63.8	58	0	0.0	3
RKANNØ1C	34	70.8	48	0	0.0	0
RKANCØ1J	297	98.7	301	136	98.6	138
3K Fills Pill Prescription						
AH15	276	84.4	327	115	82.1	140
AH18	266	81.8	325	106	77.4	137
BI14	223	70.1	318	82	60.7	135
5K Measures Insulin Correctly						
AH2	312	69.5	449	174	83.7	208
AH5	376	82.3	457	201	96.6	208
AH7	373	82.0	455	202	96.7	209
AH8	412	89.4	461	206	98.6	209
BI1	376	97.9	384	176	100.0	176
BI5	253	69.5	364	140	80.9	173
BI7	310	80.7	384	173	98.3	176
99K Insulin at Correct Times						
AH6	391	84.8	461	200	95.7	209
AH11	269	59.4	453	153	74.3	206
BI3	261	68.3	382	146	83.0	176
BI6	339	88.7	382	174	98.9	176
RKAN%99Q	262	86.8	302	166	86.5	192
RKAND99R	242	93.4	259	165	95.4	173

Table 12 continued

Delphi Groupings/Items	Total Sample f	%	N	Analysis Sample f	%	N
21K Times Test Correctly						
BI26	125	59.2	211	73	68.2	107
BI30	185	84.5	219	99	89.2	111
BI34	29	47.5	61	12	50.0	24
BI38	12	27.9	43	4	25.0	16
EKAND21C	133	49.3	270	86	62.3	138
EKAND21D	200	73.8	271	104	78.2	133
EKAND21E	27	32.5	83	13	37.1	35
EKAND21F	10	14.3	70	1	4.0	25
23K Stores Test Materials Properly						
AH19	422	90.2	468	192	93.7	205
AH20	432	93.7	461	192	94.6	203
BI21	375	96.9	387	174	98.3	177
26K Tests Urine before Meals						
AH21	392	86.0	456	192	94.1	204
AH22	414	88.8	466	189	92.2	205
BI20	324	83.9	386	171	97.2	176
BI22	329	84.4	390	153	86.4	177
28K Interprets Test Correctly						
AH25	281	96.6	291	136	97.8	139
AH26	185	64.9	285	105	75.5	139
AH29	247	88.2	280	117	87.5	134
AH30	253	89.7	282	123	91.8	134
AH33	63	71.6	88	26	72.2	36
AH34	74	81.3	91	30	83.2	36
AH37	41	56.2	73	14	53.8	26
AH38	49	67.1	73	15	57.7	26
BI27	158	74.5	212	81	75.7	107
BI28	199	93.4	213	107	100.0	107
BI31	199	93.4	213	103	95.4	108
BI32	166	78.3	212	94	86.2	109
BI35	35	61.4	57	17	77.3	22
BI36	44	75.9	58	17	73.9	23
BI39	11	28.9	38	4	25.0	16
BI40	26	61.9	42	12	75.0	16
RKAAD28J	169	59.5	284	101	66.9	151
33K Eats Prescribed Meals at Prescribed Times						
AH39	476	98.8	482	208	99.5	209
AH40	402	86.6	464	191	91.4	209
AH42	450	94.5	476	202	96.7	209
AH43	283	61.4	461	147	71.0	207
BI41	373	96.6	386	175	98.9	177

Table 12 continued

Delphi Groupings/Items	Total Sample f	%	N	Analysis Sample f	%	N
33K Eats Prescribed Meals at Prescribed Times (cont.)						
BI42	358	93.5	383	173	98.3	176
RKANC33M	236	86.4	273	141	94.0	150
RKANC33N	245	90.4	271	146	97.3	150
RKANC33E	223	82.3	271	128	85.3	150
41K Uses Food Exchange List						
AH44	264	56.5	467	131	63.0	208
AH46	224	56.3	398	116	63.7	182
AH47	384	87.1	441	183	93.4	196
BI45	201	55.7	361	115	66.9	172
BI46	264	73.1	361	139	80.8	172
DKAAN41M	33	27.5	120	31	36.5	85
DKAAN41V	29	27.6	105	28	36.8	76
DKAAN41F	26	22.8	114	24	29.6	81
DKAAN41C	47	40.5	116	46	52.9	87
DKAAN41P	53	42.7	124	52	58.4	89
DKAAN41T	43	32.6	132	42	44.7	94
42K Plans Alcohol Intake						
AH45	402	84.8	474	175	84.1	208
AH48	128	30.6	418	66	34.4	192
BI44	266	71.9	370	131	77.1	170
44K Weighs/Measures Food						
AH49	330	76.6	431	157	81.3	193
BI43	361	94.8	381	171	97.2	176
DKANCWGM	203	61.9	328	112	63.6	176
DKANCFDL	231	72.0	321	137	80.1	171
DKBNCFDL	200	71.4	280	117	80.1	146
RKANNCAL	97	39.0	249	60	41.1	146
37K Achieves/Maintains Correct Weight						
AH41	470	98.9	475	204	98.6	207
DKANNWGT	72	24.7	291	40	26.5	151

Table 12 continued

Delphi Groupings/Items	Total Sample f	%	N	Analysis Sample f	%	N
54K Inspects Feet Daily						
BI15Z1R	359	90.9	395	165	93.2	177
BI15Z2R	299	75.7	395	141	79.7	177
BI15Z3R	287	72.7	395	136	76.8	177
BI15Z4R	358	90.6	395	161	91.0	177
BI15Z5R	374	94.7	395	168	94.9	177

[a] Delphi Item grouping number (see Appendix A to Summary Report).

[b] Question number (see Table III-A-1).

[c] Name of variable created from two or more items, usually the patient's answer and information from the Medical Records Abstraction Form (see Table III-A-2).

Table 13

TEST-RETEST RELIABILITY COEFFICIENTS FOR 21 SINGLE-ITEM KNOWLEDGE MEASURES INCLUDED IN FORMS A AND B

Global Concept/Items[a]	Delphi Group	N	Test-Retest Coefficient
Safety Self-Care (SAFSLF)			
AH12ZB-BI11Z2	34K	363	.40
AH14ZB-BI12	34K	337	.67
AH12ZG-BI11Z5	36K	351	-.33
AH13ZC-BI10	36K	360	.25
AH12ZD-BI11Z3	36K	163	-.16
AH4-BI4	61K	366	.08
AH10-BI9	61K	364	.39
AH9-BI8	62K	339	.45
Oral Medication (MEDOR)			
AH17-BI13	1K	243	.62
AH15-BI14	3K	249	.44
Insulin Taking (MEDINS)			
AH7-BI7	5K	360	.71
AH8-BI5	5K	353	.42
AH11-BI3	99K	364	.70
Urine Testing-Technical (URTEC)			
EKAND21C-BI26	21K	176	.70
EKAND21D-BI30	21K	192	.49
EKAND21E-BI34	21K	46	.68
EKAND21F-BI38	21K	27	.57
AH21-BI20	26K	359	.62
AH25-BI27	28K	188	.25
AH26-BI28	28K	186	.21
Nutrition-Technical (NUTTEC)			
DKANCFDL-DKBNCFDL	44K	276	.76

[a]Form A (first questionnaire) and Form B (retest questionnaire) variable names are separated by hyphens. Consult Tables III-A-1 and III-A-2 for definitions of the items.

All of the very low test-retest coefficients involved items having very skewed score distributions (i.e., about 85 to 98 percent correct responses).

Testing Hypothesized Knowledge Subtests

Twenty-one multi-item knowledge subtests were hypothesized. Two knowledge areas (11K, 59K) represented by only one item are not included in this total. Also, one of the 21 subtests (from Form B) was excluded from the scaling evaluation due to lack of response variance. The scaling evaluations were performed in two rounds of analyses. The first round tested 20 multi-item knowledge subtests exactly as hypothesized. The second round included four subtests that were modified on the basis of results from the first round. For example, some very poor items were eliminated after the first round. The first two rounds of analyses were performed for respondents having little or no missing data. Following the second round of analyses, i.e., after scaling decisions had been made, descriptive statistics and reliability estimates were computed using all available data; the previously explained rules for estimating missing responses were followed.

It was necessary to exclude (from the first round of analyses) three multi-item knowledge subtests (27K, 53K, 54K) constructed from Form B items. The ANLITH program used to evaluate subtests in terms of scaling criteria is limited to a matrix containing 19 subtests. Following elimination of two Form A subtests, which failed all scaling criteria, it was possible to add the Form B subtests to the multitrain matrix generated by ANLITH during the second round of analyses.

The first scaling criterion (convergent validity) required that each knowledge item correlate substantially (0.30 or higher) with the total score for the subtest it was hypothesized to measure. Because many subtests contained only a few items, item-test correlations were corrected for overlap to provide a more easily interpretable test. In other words, each item was correlated with the sum of other items in the same subtest. Thirteen of the 18 multi-item knowledge subtests constructed from Form A satisfied this criterion (data not reported). Five of those subtests (58K, 62K, 21K, 26K, 37K) did not. Of the three

Form B multi-item knowledge subtests, one (54K) passed the first criterion; one (27K) failed, and the third (53K) could not be tested due to lack of item score variability.

The second scaling criterion (discriminant validity) required that each knowledge item correlated higher with its hypothesized subtest than with other subtests in the same matrix. A summary of results for this criterion applied to 18 knowledge subtests is presented in Table 14. These 18 subtests were hypothesized from Form A Delphi item groupings. Several important trends are apparent in Table II-4; other trends identified during the review of the item-test correlation matrix (data not reported) and preparation of the summary will also be noted.

Attempts to construct knowledge subtests from Delphi item groupings were successful in many instances. Out of 1224 tests of the discriminant validity criterion, 865 (approximately 71 percent) supported item validity. Further, the majority of scaling errors were probable (as opposed to definite). This overall pattern of results should be viewed favorably given that this was the first field test of a new instrument designed to measure numerous knowledge constructs.

Table 14 also indicates clearly that scaling errors were concentrated in four of the five subtests (e.g., 21K, 37K, 58K, and 52K) that failed the first scaling criterion. Scaling errors involving the short two-item subtests were traced to very low (and sometimes) negative correlations between subtest items (these subtests will prove very unreliable by internal-consistency standards). In other instances including some of the longer subtests (e.g., 33K and 41K), scaling errors were caused by a few poor items that should be eliminated or revised. This pattern of results suggests that revisions in subtests based on these data (e.g., item deletions) are likely to produce noteworthy improvements in scaling results and reliability estimates. A third source of scaling errors was substantial overlap between some scales (e.g., 34K and 36K, which were both hypothesized to measure knowledge of self-care safety behaviors). Combination of very highly overlapping scales is likely to reduce scaling errors with little or no loss of information. This strategy for improving

Table 14

SUMMARY OF DISCRIMINANT VALIDITY EVALUATIONS FOR ITEMS IN 18 FORM A KNOWLEDGE SUBTESTS BASED ON DELPHI GROUPINGS

Global Concepts and Subtests[a]	Number of Items	Success Ratio	Errors	
			Probable	Definite
Safety Self-Care				
34K Changes Eating Time[c]	5	81/85	4	0
36K Eats for Hypoglycemia[c]	11	156/187	29	2
58K Carries Emergency Glucose	2	5/34	17	12
61K Carries Diabetic ID[c]	3	44/51	5	2
62K Tells about Hypoglycemia	2	4/34	14	16
Oral Medication				
1K Knows Pill Dosage[c]	2	32/34	2	0
3K Fills Pill Prescription[c]	2	32/34	2	0
Insulin Taking				
5K Measures Insulin Correctly[c]	4	67/68	1	0
99K Insulin at Correct Times	4	33/68	17	18
Urine Testing				
21K Times Test Correctly	4	0/68	3	65
23K Stores Test Materials Properly[c]	2	32/34	2	0
26K Tests Urine before Meals[c]	2	32/34	2	0
28K Interprets Test Correctly[c]	5	55/85	20	10
Nutrition--Eating				
33K Eats Prescribed Meals	7	68/119	34	17
Nutrition--Technical				
41K Uses Food Exchange List[c]	9	140/153	11	2
42K Plans Alcohol Intake[c]	2	24/34	10	0
44K Weighs/Measures Food[c]	4	59/68	8	1
Nutrition--Body Weight				
37K Achieves/Maintains Correct Weight	2	1/34	22	11
TOTALS	72	865/1224	203	156

[a] Three Form B multi-item knowledge subtests (27K, 53K, 54K) were not included in these analyses. Also, two Delphi groupings (11K, 59K) represented by only one item each were not included.

[b] Number of definite scaling successes over the number of tests.

[c] Satisfied discriminant validity tests.

knowledge subtests is explored in a subsection to follow (see Construction of General Knowledge Subtests).

Revised Knowledge Subtests

Four of the knowledge subtests (36K, 99K, 28K, 33K) contained poor items that accounted for a disproportionately large number of scaling errors in the discriminant validity evaluations and that substantially lowered subtest homogeneity (i.e., did not correlate consistently well with other items in the same subtest). These items were discarded and the discriminant validity criterion was applied to the revised subtests. Results are summarized in Table 15. All four revised subtests were substantially improved. The success ratios for discriminant validity evaluations increased from 6 to 42 percentage points for the four revised subtests. It is recommended that these revisions be adopted in future field tests.

Table 15

SUMMARY OF DISCRIMINANT VALIDITY EVALUATIONS FOR ITEMS IN FOUR REVISED KNOWLEDGE SUBTESTS

Subtests	Number of Items	Success[a] Ratio	Errors	
			Probable	Definite
36K Eats for Hypoglycemia	10	152/170	18	0
99K Takes Insulin at Correct Times	2	34/34	0	0
28K Interprets Test Correctly	3	49/51	2	0
33K Eats Prescribed Meals	5	61/85	18	6

[a]Number of definite scaling successes over number of tests.

Reliability of Subtest Scores

Internal-consistency and/or test-retest reliability coefficients were computed for 20 knowledge subtests constructed from two or more items each. All internal-consistency estimates were computed from correlations among Form A items except in two instances (53K and 54K where multi-item subtests were fielded only in Form B (the retest questionnaire). One two-item subtest (27K) was not included due to lack of item response variability; also, two knowledge areas (11K, 59K) represented by only one item each were not included. Test-retest reliability was estimated for multi-item scales fielded in Forms A and B, regardless of whether the exact items were fielded both times. If the 2 scales contained different numbers of items or differently worded items, the retest coefficient should be interpreted as a conservative estimate of reliability. Test-retest estimates were computed only for those subtests that contained at least two items from the Form A (test) and Form B (retest) questionnaires.

Table 16 presents a summary of information about the number items used to score Form A and B versions of 20 knowledge subtests, the subtest homogeneity (average inter-item correlations) and internal-consistency and test-retest estimates of reliability. Homogeneity coefficients facilitate comparisons between tests differeing in length.

Two Form B knowledge subtests (53K and 54K) not inlcuded in the Form A scaling studies described previously were added to the study of subtest reliability. Revisions in four subtests suggested by the scaling studies were implemented before estimating reliability. Internal-consistency reliability estimates for the 20 subtests ranged from zero to 0.90p the median coefficient was approximately 0.47. Test-retest estimates for 10 of these subtests ranged from zero to 0.84 and the median test-retest coefficient was approximately 0.47.

Eleven of the 14 subtests that satisfied convergent and discriminant criteria during the scaling evaluation achieved the minimum standard for internal-consistency reliability (0.50 or higher) or came very close (within a few points) that standard. There were no subtests that met the internal consistency standard without passing the scaling evaluations.

Table 16

INTERNAL-CONSISTENCY AND TEST-RETEST RELIABILITY ESTIMATES FOR 20 MULTI-ITEM KNOWLEDGE SUBTESTS

Global Concepts and Subtests	Number of Items A	Number of Items B	Homogeneity	Reliability: Internal Consistency	Reliability: Test-Retest
Safety Self-Care					
34K Changes Eating Time	5	1	.36	.74	(a)
36K Eats for Hypoglycemia[b]	10	6	.19	.70	.51
58K Carries Emergency Glucose	2	1	.05	.09	--
61K Carries Diabetic ID	3	2	.18	.40	.00
62K Tells about Hypoglycemia	2	1	.05	.09	--
Safety MD Care					
53K Reports Symptoms to MD	0	3	.08[c]	.24[c]	--
Oral Medication					
1K Knows Pill Dosage	5	1	.70	.83	--
3K Fills Pill Prescription	2	1	.82	.90	--
Insulin Taking					
5K Measures Insulin Correctly	4	3	.25	.57	.44
99K Takes Insulin at Correct Times[b]	2	2	.35	.52	.34
Urine Testing					
21K Times Test Correctly	2[d]	4	.00	.00	.84[e]
23K Stores Test Materials Properly	2	1	.32	.48	--
26K Tests Urine before Meals	2	2	.26	.41	.19
28K Interprets Test Correctly[b]	3	8	.19	.42	.66[e]
Nutrition--Eating					
33K Prescribed Meals[b]	5	2	.38	.75	.11
Nutrition-Technical					
41K Uses Food Exchange List	9	2	.27	.75	.31
42K Plans Alcohol Intake	2	1	.21	.35	--
44K Weighs/Measures Food	4	2	.18	.47	.50
Nutrition--Body Weight					
37K Achieves/Maintains Correct Weight	2	0	.05	.10	--
Foot Care					
54K Inspects Feet Daily	0	5	.27[c]	.65[c]	

[a] Not computed due to lack of multi-item scale at test or retest.

[b] Revised after the first round of scaling evaluations (see text).

[c] Based on Form B respondents; all other homogeneity and internal-consistency reliability estimates based on Form A (N=209).

[d] Two items eliminated due to lack of response variance.

[e] Small sample sizes (N=7-14) were encountered for two tests; all other test-retest coefficients were computed from the responses of 175-209 respondents.

Means and standard deviations for 20 Form A and Form B knowledge subtests are presented in Table 17. These estimates were based on the smaller analysis sample (N=209).

Construction of General Knowledge Tests

A principal components analysis suggested that patient knowledge would be described in terms of a general knowledge factor and possibly higher order knowledge dimensions. A noteworthy general knowledge factor was identified; it accounted for approximately one-fifth of the total measured variance (included error) in the 20 Form A subtests and for approximately 45 percent of the estimated total reliable variance. A Scree Test and other rotational criteria supported interpretation of four principal components associated with eigenvalues greater than unity. These components were rotated to orthogonal simple structure to facilitate interpretation (data not reported).

Interpretation of the four components was both difficult and straightforward and was both consistent and inconsistent with hypothesized global knowledge concepts. Completely consistent with global hypotheses were factors defined by subtests in the oral medication and urine testing categories. The following subtests correlated significantly with the oral medication component (factor loadings shown in parentheses): 1K (0.89) and 3K (0.88). Refer to Table 17 for definitions of these global concepts and for their subtests.) These subtests did not correlate significantly (0.30) or higher with any other factors. The following subtests correlated significantly with the urine testing component: 21K (0.71), 28K (0.68), 26K (0.44), and 23K (0.35). With the exception of the latter subtest, these measures did not correlate significantly with any other components.

The two additional general knowledge factors involved subtests from more than one general knowledge concept. One component was composed of knowledge subtests having to do with eating (44K and 58K) and diabetic identification (61K). Loadings of .064, 0.56, and 0.65 were observed for these subtests, respectively. The remaining component and largest (in terms of variance accounted for) contained significant loadings for subtests from the following areas: self-care

Table 17

MEANS AND STANDARD DEVIATIONS FOR 20 KNOWLEDGE SUBTESTS
(N=209)

Global Concepts/Subtests	Number of Items	Mean	SD
Safety Self-Care			
34K Changes Eating Time	5	3.58	1.40
36K Eats for Hypoglycemia[a]	11	8.42	2.10
58K Carries Emergency Glucose	2	1.86	.36
61K Carries Diabetic ID	3	2.82	.47
62K Tells about Hypoglycemia	2	1.69	.50
Safety MD Care			
53K Reports Symptoms to MD[b]	3	2.88	.99
Oral Medication			
1K Knows Pill Dosage	5	1.58	1.15
3K Fills Pill Prescription	2	1.06	.95
Insulin Taking			
5K Measures Insulin Correctly	4	3.75	.63
99K Insulin at Correct Times[a]	4	3.27	.85
Urine Testing			
21K Times Test Correctly	4	.98	.70
23K Stores Test Materials Properly	2	1.93	.49
26K Tests Urine before Meals	2	1.82	.45
28K Interprets Test Correctly[a]	9	3.19	1.52
Nutrition--Eating			
33K Eats Prescribed Meals[a]	7	5.56	1.55
Nutrition--Technical			
41K Uses Food Exchange List	9	3.12	2.16
42K Plans Alcohol Intake	2	1.15	.65
44K Weighs/Measures Food	4	2.23	1.16
Nutrition--Body Weight			
37K Achieves/Maintains Correct Weight	2	1.17	.43
Foot Care			
54K Inspects Feet Daily[a]	5	4.36	1.01

[a] Scored from Form B items; all other subtests constructed from Form A.
[b] Revised after the first round of scaling evaluations (see text).

safety (34K, 36K, and 62K), safety MD care (53K), and insulin taking (5K, 99K). All of these subtests correlated highly with the component (0.76 to 0.47) and did not correlate significantly with any other components. While the emphasis of this component is on safety-related knowledge, the presence of the latter two subtests suggests a more general interpretation.

Four general knowledge tests were constructed using items in the subtests identified above. Ad hoc construction of these tests virtually eliminated item discriminant validity errors and internal consistency coefficients for these tests ranged from 0.65 to 0.90. Table 18 presents a summary of information about the four general knowledge tests including descriptive statistics and internal-consistency reliability estimates based on the larger analysis sample (N=310).

Correlation with Education and Guessing Response Bias

The four general knowledge tests and 20 subtests were correlated with education to enhance understanding of their validity in relation to the hypothesis that knowledge increases with formal education. Measures of two kinds of response bias (guessing, and favorable compliance response bias) were also correlated with these tests to increase understanding of their validity and how response styles relate to knowledge.

Table 19 presents the results of these analyses (N=209). Eleven of the 24 correlations between education and knowledge tests were statistically significant ($p < 0.01$). All significant correlations were in the positive direction, as hypothesized, indicating increasing knowledge with education. Correlations between self-reported guessing and knowledge tests were neither statistically significant nor very large. Three of 24 for compliance bias correlations were significant; all were negative. These trends suggest that tendencies to guess do not bias knowledge test scores and that compliance bias tendencies may be greater for less informed respondents or that knowledge test scores contain some systematic bias. The former explanation is more plausible than the latter.

Table 18

DESCRIPTIVE STATISTICS AND RELIABILITY ESTIMATES FOR FOUR GENERAL KNOWLEDGE TESTS (N=310)

Tests[a]	Number of Items[b]	Mean	SD	Reliability[c] Estimate
Eating		6.91	1.72	.65
General Safety		12.34	4.56	.90
Oral Medication		3.17	1.30	.82
Urine Testing		9.11	2.06	.79

[a]Tentative labels.

[b]Also defines highest possible score.

[c]Internal-consistency method.

VALIDITY

This section presents findings relevant to the validity of some of the specific knowledge measures described previously. Analyses were conducted on a group of diabetic patients who completed a knowledge questionnaire and demonstrated some of their self-care activities.

The results do not support the validity of the five scales evaluated, but the small sample sizes and low amounts of score variation limit the power of the tests to detect validity. Nevertheless, the findings raise important questions for future research: What kinds of knowledge are necessary for successful self-care performance? To what extent is knowledge sufficient to predict performance?

Table 19

CORRELATIONS OF KNOWLEDGE SCALES WITH EDUCATION LEVEL, SELF-REPORTED GUESSING AND THE COMPLIANCE RESPONSE BIAS INDEX

Scales	Education	Guessing	Compliance Response Bias
General Tests			
Eating	-.12	-.02	.03
General Safety	.42*	.05	-.16
Oral Medication	.10	.02	-.09
Urine Testing	.20	-.06	-.05
Subtests			
36K	.36*	.04	-.13
61K	.03	.02	.01
05K	.27*	.05	-.22*
21K	.22	-.06	-.02
26K	-.12	.08	.05
28K	.21*	-.06	-.01
33K	.29	.14	-.22*
41K	.07	.00	-.02
44K	-.15	.02	.02
99K	.08	.03	-.12
34K	.39*	.05	-.13
58K	-.04	.01	.04
62K	.29*	.03	-.18*
01K	.10	.02	-.10
03K	.10	.02	-.07
23K	.00	-.07	-.02
42K	.27*	-.04	-.05
37K	.19*	-.02*	-.02
27K	-.03*	.14	.02
53K	.16*	.07	-.01*
54K	.21	-.03	-.13

*$p \leq .05$

Methodology

Sample Selection. The sample consisted of 85 patients who had arrived at one of two clinics for a doctor's appointment. The patients were asked to participate in a demonstration of procedures involved in urine sugar tests, acetone tests, and insulin measurement. All patients were insulin-taking, English speaking adults. Before the demonstration, the patients were asked to complete a short knowledge battery of of tests if they had not filled out the diabetes questionnaires previously.

Demonstration Procedures and Scores. As part of the demonstration, patients were asked to perform several tasks. First, they were asked to conduct three urine tests using the solutions provided by the interviewer. Patients were asked to conduct each test the way they would do them at home and then write the test results in the appropriate columns of the questionnaire. Two of the tests were sugar tests and the third was an acetone test. For the sugar tests, patients were given one solution of distilled water containing two percent glucose and one solution of tinted distilled water containing no glucose. The order of solutions presented was counterbalanced across subjects. As the patients performed each test, the interviewer recorded data relevant to the procedures demonstrated. At the end of each test, the patient was asked to record the test results and describe how much sugar was contained in each solution.

For the acetone test, patients were given one solution containing water and a moderate percentage of fingernail polish remover. Again, patients were told to conduct the test just like they would do it at home. The interviewer noted number of tablets used, number of drops of urine put on the tablet, number of seconds the patient waited before making final comparison to the color chart, and unusual or incorrect procedures. At the end of the test, the patient was asked to record the test result and describe the amount of acetone in the solution.

The second part of the demonstration asked patients to prepare an insulin dosage as they did each morning. Patients were shown a number of insulin bottles and syringes and asked to select the type(s) and strength of insulin as well as the syringe normally used. As the patient prepared his insulin dosage, the interviewer recorded the number of units withdrawn and noted incorrect or unusual procedures.

The third section of the demonstration required the patient to show anything he had with him to eat in the event of an insulin reaction and to show any types of identification with him that said he was a diabetic.

Five scores indicating performance quality were constructed for each patient:

1) The number of seconds waited during the sugar tests before making final comparisons to the color chart.
2) The number of correct interpretations of the amount of sugar in the sugar test solutions.
3) The amount of insulin withdrawn during the insulin test.
4) The type of food the patient brought with him for use during an insulin reaction.
5) The type of identification the patient had that said he was a diabetic.

The patient's response to each of the demonstration items was scored as correct or incorrect (e.g., he waited the correct number of seconds and carried something with sugar in it), summed (if necessary), and the total score was compared to the knowledge scale relevant to each specific construct.

Knowledge Measures. Five knowledge scale scores were constructed for each patient according to the rules developed for the total sample. These five scores pertained to five item groupings: (21K) times test correctly, (28K) interprets test correctly, (61K) carries diabetic identification, (5K) measures insulin correctly, and (58K) carries emergency glucose.

The Time Test Correctly scale contained four items about knowledge of the number of seconds a patient is supposed to wait before comparing the color of the sugar test solution to the color chart. Scores were dichotomized as correct or incorrect. Since patients responded to only one of the four items (depending on which type of sugar test was used), the final score was a single-item measure.

The Interprets Test Correctly scale reflected responses to eight items pertaining to knowledge of interpretation of colors and test readings. Again, patients responded to only two of the eight items depending on which type of test used. Scores ranged from 0 to 2 with 0 indicating incorrect knowledge, 1 indicating correct knowledge of only one item, and 2 indicating correct knowledge of both items.

The Carries Diabetic Identification scale contained two items pertaining to knowledge of the importance of carrying a diabetic identification. Scores ranged from 0 to 2 with 0 indicating incorrect knowledge of both items, 1 indicating correct knowledge of only one item, and 2 indicating correct knowledge of both items. This scale contained one less item than than developed for the total sample described previously.

The Measures Insulin Correctly scale contained four items pertaining to knowledge of the type of syringes used with two kinds of insulin, the measurement units of insulin dosage, and the importance of measuring insulin before it is injected. Scores ranged from 0 to 4 with 0 indicating incorrect knowledge of all items.

The Carries Emergency Glucose score was based on response to only one item pertaining to knowledge that diabetics who take insulin should carry something with sugar in it in case they have an insulin reaction. Scores were dichotomized as correct (1) and incorrect (0). This scale contains only one item whereas the scale developed for the main patient sample contained two items. For all five scales, patients responding "don't know" to an item were scored as responding incorrectly to that item.

Results.

The logic of the knowledge scale validity analysis can be illustrated with the 2x2 table:

		Skill: Low	Skill: High
Knowledge	Low	A Valid	B Invalid
	High	C Predictive Validity Information	D Valid

A knowledge variable is completely valid if all of the observations fall into cells A and D. It is not completely valid if the number of observations in cell B exceeds an expected chance frequency. Observations in cell C have different interpretations depending on the kind of validity addressed: They detract from the predictive validity of a measure but they are irrelevant to the intrinsic validity of concern here. The latter follows from the assumption that knowledge is a necessary but insufficient condition for skill. The cell C cases represent people who may have requisite knowledge but do not exhibit skill. So a high frequency of cell C cases does not mean that the knowledge measure is not measuring knowledge.

Applying this logic to the five cross-tabulations in Table 20 interpretations of the validity of specific measures can be made. The knowledge measure regarding timing of sugar tests appears to have poor validity. Too many subjects fall into the low knowledge/high skill cell and are thus misclassified by the knowledge scale.

The knowledge measure regarding interpretation of sugar tests also appears to be invalid because there is variation in the knowledge scores but almost none in the criterion variable.

Interpretation of knowledge pertaining to insulin measurement is less clearcut. The validity of the measure depends upon how rigidly one defines adequate knowledge. If correct scores on two of the four items is used as a cutoff, then measured knowledge does appear to be a necessary condition for skill. If knowledge is defined as answering all four questions correctly, then the measure appears invalid (too many subjects fall into the low knowledge/high skill cells).

Knowledge of the importance of carrying glucose has no variance. Almost all subjects know that they should carry emergency glucose, but they do not necessarily carry it. Since the demonstration is really a measure of behavior compliance, we also learn that knowledge is not necessarily related to this kind of compliance.

Similarly, most patients know the importance of carrying a diabetic identification. But compliance (actually carrying an identification with them) is not highly related to the knowledge. Also, the predictive validity of this knowledge measure is low.

Table 20

CROSS-TABULATIONS BETWEEN KNOWLEDGE MEASURES AND BEHAVIORS (SKILLS)

Knowledge: Knows how to time sugar tests correctly

Demonstration: Times test correctly in one or more tests.

	0	1 or 2
0	25	23
1	8	15

Knowledge: Knows how to interpret sugar test correctly

Demonstration: Interprets sugar test correctly in one or more tests.

	0	1 or 2
0	0	1
1	0	24
2	2	47

Knowledge: Knows how to measure insulin correctly.

Demonstration: Withdraws correct units of insulin.

	0	1
1	0	1
2	1	5
3	4	21
4	7	42

Knowledge: Knows about carrying emergency glucose

Demonstration: Carries something with sugar in it.

	0	1
0	1	0
1	42	38

Knowledge: Knows how to measure insulin correctly.

Demonstration: Carries Diabetic ID

	0	1
0	1	1
1	2	7
2	21	45

These findings do not support the validity of any of the knowledge measurement scales. Insulin measurement knowledge may be internally valid depending on how rigidly one defines knowledge, but its predictive validity is low.

Further research needs to be conducted before any definitive statements can be made about the validity of knowledge measures. The findings suggest that some of the scales contain items measuring knowledge that is not necessary for adequate performance. Results also suggest that high knowledge is not always sufficient for good performance, implying that programs aimed only at education may not change skill or behavioral levels. It should be kept in mind, though, that the findings were limited by small sample sizes and the small subset of possible measures which were tested.

These preliminary analyses emphasize the importance of future efforts to define what knowledge is necessary for successful performance and to elucidate the relationship between knowledge and behavioral skill. Such efforts will have substantial implications for the design of educational programs designed to produce changes in behavior.

ANNEX TO SECTION III

Table III-A-1

DEFINITIONS AND CORRECT ANSWERS FOR 107 KNOWLEDGE VARIABLES NOT REQUIRING RECORD CHECKS

Label/Form A	Label/Form B	Definition	Correct Answer
AH14ZA		Increasing insulin before exercise prevents insulin reaction	No
AH14ZB		Eating snack before exercise prevents insulin reaction	Yes
AH14ZC		Cutting down on food if exercising prevents insulin reaction	No
AH14ZD		Drinking coffee after exercising prevents insulin reaction	No
	BI12	Diabetic planning to exercise should	Eat extra before exercising
AH12ZA		Fruity smell to breath is symptom of insulin reaction	No
AH12ZB		Trembling is symptom of insulin reaction	Yes
AH12ZC		Cold skin is symptom of insulin reaction	Yes
AH12ZD		Stomach cramps is symptom of insulin reaction	No
AH12ZE		Feeling hungry is symptom of insulin reaction	Yes
AH12ZF		Blurred vision is symptom of insulin reaction	Yes
AH12ZG		Feeling thirsty is symptom of insulin reaction	No
AH13ZA		Get exercise when early signs of insulin reaction	No
AH13ZB		Take nap when early signs of insulin reaction	No
AH13ZC		Eat something with sugar when early signs of insulin reaction	Yes
AH13ZD		Take pills when early signs of insulin reaction	No
	BI10	What diabetic should do at first signs of insulin reaction	Eat something with sugar in it
	BI11Z1R	Urinating a lot common symptom of insulin reaction	No
	BI11Z3R	Trembling common symptom of insulin reaction	Yes
	BI11Z5R	Bad stomach cramps common symptom of insulin reaction	No
	BI11Z2R	A cold sweat common symptom of insulin reation	Yes
	BI11Z4R	Feeling very thirsty common symptom of insulin reaction	No

AA13		Supposed to have something with sugar in it	Yes
AH1		Diabetics should carry sugar	True
	BI2	Sugarless gum is good to carry for insulin reaction	False
AA11		Supposed to wear ID	Yes
AH4		Not necessary to carry ID	False
AH10		When should ID be carried	At all times
	BI4	Diabetics should wear ID	True
	BI9	When should ID be carried	At all times
AH3		Should tell others symptoms of hypoglycemia	True
AH9		Most important thing to tell families	How to take care of patient during insulin reaction
	BI8	Most important thing to tell families	How to take care of patient during insulin reaction
	BI17	Tell doctor if urine test shows high sugar and high acetone	Yes, right away
	BI18	Tell doctor if can't get right amount of exercise in one day	No
	BI19	Tell doctor if blurred vision more than one week	Yes, right away
	BI16	Tell doctor of fainting	Yes, right away
AH16		Take more pills to make up for missed one	No
AH17		When should pills be taken	Every day
	BI13	When should pills be taken	Every day
AH15		When should pill prescription be refilled	Before run out of pills
AH18		Fill pill prescription right away	Yes
	BI14	When should pill prescription be refilled	Before run out of pills
AH2		Use U-40 syringe with U-80 insulin	False
AH5		No need to measure U-100 insulin	False
AH7		Use what type syringe with U-100 insulin	U-100 syringe
AH8		Insulin dosage measured in	Units
	BI1	Insulin should be carefully measured	True
	BI5	Insulin measured in millimeters	False
	BI7	Use U-100 syringe with U-100 insulin	True
AH6		If told to take 10 units insulin in morning, OK to take 5 units bedtime and 5 units next morning	False
AH11		When ill and unable to eat you should	Continue to take insulin
	BI3	When ill and unable to eat stop taking insulin	False
	BI6	May inject insulin any time of day	False
	BI26	Time wait to compare color Clinitest	15 seconds
	BI30	Time wait to compare color Diastix test	30 seconds
	BI34	Time wait to compare color Testape	1 minute
	BI38	Time wait to compare color Clinistix	10 seconds
EKAND21C		Time wait to compare color Clinitest	14, 15, 16 seconds
EKAND21D		Time wait to compare color Diastix test	29, 30, 31 seconds
EKAND21E[a]		Time wait to compare color Testape	59, 60, 61 seconds
EKAND21F[a]		Time wait to compare color Clinistix	9, 10, 11 seconds
AH19		Urine test materials stored in refrigerator	No
AH20		Good place to store urine test materials	Cool, dry place

	BI21	Store urine test materials in dry place	Yes
AH21		When should tests for sugar be done	Just before meals
AH22		Test urine after snack	No
	BI20	When do tests for sugar	Just before meals
	BI22	When should tests for sugar be done a day	Before breakfast
AH25		Blue Clinitest solution means	No sugar in urine
AH26		Clinitest reading of 4+ means a little sugar	No
AH29		Olive green Diastix test means some sugar	Yes
AH30[a]		Diastix reading of 4+ means	A lot of sugar
AH33[a]		Light green Testape means	A little sugar
AH34[a]		Testape reading of 2+ means some sugar	Yes
AH37[a]		Dark purple Clinistix means a lot of sugar	Yes
AH38[a]		Clinistix reading of 1+ means	A little sugar
	BI27	Blue Clinitest means a little sugar	No
	BI28	Clinitest reading of 4+ means	A lot of sugar
	BI31	Brown Diastix test means	A lot of sugar
	BI32	Diastix reading of 1+ means a lot of sugar	No
	BI35	Dark blue Testape means a little sugar in urine	No
	BI36	Testape reading of 1+ means	A little sugar
	BI39	Pink Clinistix means	No sugar
	BI40	Clinistix reading of 3+ means no sugar	No
AH39		Diabetics should eat same time every day	True
AH40		If meal is skipped eat twice as much at next meal	False
AH42		East as much "diatetic" food as desired	False
AH43		Diabetic diet is	A well-balanced diet that the whole family can use
	BI41	Diabetics should eat at least three meals a day	True
	BI42	If right amount of calories eaten doesn't matter what you eat	False
AH44		May eat as much of which foods	clear broth
AH46		One orange exchanged for	1 small apple
AH47		One slice bread exchanged for one small potato	Yes
	BI45	1 egg exchanged for	1 slice cheese
	BI46	2 bread exchanges and 1 meat exchange means	Eat 2 items on bread list and 1 item on meat list
AH45		Should count calories in cocktail when planning meal	Yes
AH48		One bottle of beer	Equals 1 bread and 2 fat exchanges
	BI44	What statement about alcohol is true	Must be counted in daily meal plan
AH49		What best measures food	Measuring cups and spoons
	BI43	Should weigh or measure food	True

AH41		Weight control is important	True
	BI15Z2R	Good foot care: inspect feel daily	Yes
	BI15Z4R	Good foot care: keep toenails short	No
	BI15Z5R	Good foot care: cut toenails straight	Yes
	BI15Z1R	Good foot care: cut corns regularly	No
	BI15Z3R	Good foot care: walk barefoot	No

---

[a] Deleted from analysis due to small sample size or lack of score variability.

Table III-A-2

DEFINITIONS OF 23 KNOWLEDGE VARIABLES
SCORED BY USING MEDICAL RECORD INFORMATION

Label/Form A	Label/Form B	Definition
RKANC34A		Knows whether to adjust meal/snack schedule
DKANC11A		Knows whether to adjust medication
RKAND01B [a]		Knows what pill strength to take
RKANN01C [a]		Knows prescribed number of pills
RKANC01J [a]		Knows whether to take oral medication
RKAN%99Q		Knows kind of insulin prescribed
RKAND99R		Knows insulin strength to use
RKAAD28J		Knows correct acetone test
RKANC33M		Knows whether to eat morning snack
RKANC33N		Knows whether to eat afternoon snack
RKANC33E		Knows whether to eat evening snack
DKAAN41M		Knows amount of milk prescribed each day
DKAAN41V [a]		Knows amount of vegetables prescribed each day
DKAAN41F		Knows amount of fruit prescribed each day
DKAAN41C		Knows amount of bread/carbohydrates prescribed each day
DKAAN41P		Knows amount of meat prescribed each day
DKAAN41T		Knows amount of fat prescribed each day
DKANCWGM		Knows whether to use food exchange list or measures and weighs food
DKANCFDL		Knows whether to use food exchange list
	DKBNCFDL	Knows whether to use food exchange list
RKANNCAL		Knows prescribed daily calorie limit
DKANNWGT		Knows prescribed ideal body weight
RBAND01A [a]		Knows what kind of pills to take

[a] Deleted from analysis due to small sample size or lack of score variability.

IV. ATTITUDE MEASUREMENT

Early resource allocation decisions limited the number of attitude questions asked and the characteristics of the measurement evaluation design. Our findings about answer variability and stability are favorable and we do find support for a general attitude factor. However, we cannot recommend particular items or scales at this time.

RESPONSE VARIABILITY

A first basic result is that the tested attitude questions do obtain response variability; diabetic patients differ in the feelings and perceptions asked about. The attitude items and obtained response distirbutions are shown in Table 21. The median response category is usually in a "favorable" direction, indicating that diabetics in this sample tend to have favorable attitudes about their self-care, exceptions are in the diet/nutrition area and also about medical care costs.

RETEST STABILITY

A mid-project decision was made to eliminate the retest design for attitude items. By mistake, three attitude items did remain in the retest questionnaire so we do have some information after all.

The items were about urine testing (see Table 21) and their retest correlations were in the same range as most single behavior and knowledge item retest coefficients:

Item	Retest Correlation
AD13	.47
AD14	.69
AD15	.46

Table 21

ATTITUDE ITEMS AND ANSWER DISTRIBUTIONS

		Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree	Not Answered
AI 1.	It is hard to do urine tests just right every time.....	57	92	19	203	97	28
AI 2.	It's important for the doctor to know right away if a diabetic passes out from insulin shock (hypoglycemia).....	343	88	15	18	6	26
AI 3.	It is embarrassing for me to carry or wear identification that says I'm diabetic.....	30	62	29	171	180	24
AI 4.	It's helpful for diabetics to call the doctor if they get several high urine test readings in a row.....	214	183	29	38	9	23
AI 5.	It's important to fill your prescription for diabetes tablets..	260	109	22	2	6	97
AI 6.	Diabetics enjoy themselves less when they're out with others because they have to eat certain meals at certain times..	51	112	26	181	101	25
AI 7.	It's important for me to visit doctors and specialists regularly.....	266	170	13	18	9	20
AI 8.	If I don't take special care of my feet, I will probably have trouble with them.....	249	203	15	10	1	18
AI 9.	It's hard to know when you should report urine test results to your doctor.....	15	72	48	255	83	23
AI 10.	As long as I know what to do for insulin reactions, it's not necessary for anyone else to know.....	13	15	13	160	257	38
AI 11.	It's easy to remember to take the right amount of medicine....	179	230	18	33	12	24
AI 12.	It's a bother to have to inject insulin at just the right time.....	49	118	42	150	82	55
AI 13.	Staying on a diabetic diet is good because it keeps you trim and healthy looking.....	180	227	33	25	9	22
AI 14.	Injecting insulin is painful....	46	106	63	164	63	54
AI 15.	It's important for all of my doctors to know I have diabetes.....	354	112	2	1	3	24
AI 16.	It's easy to forget to take special care of my feet.....	40	157	22	179	76	22
AI 17.	It's hard to know when you need to report insulin reactions to your doctor.....	19	60	58	209	99	51

Table 21 continued		Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree	Not Answered
AI 18.	It's easy to remember to eat the right thing at the first sign of an insulin reaction.....	141	191	60	47	12	45
AI 19.	It's important for diabetics to keep a record of their urine test results.....	208	230	24	10	2	22
AI 20.	It costs too much to see the doctors and specialists that diabetics are told to visit.....	118	140	41	116	56	25
AI 21.	It's easy to keep my weight where it should be.....	41	139	34	170	86	26
AI 22.	It's hard to find the time to fill a prescription for diabetes tablets.....	8	10	67	172	128	111
AI 23.	Diabetics please their doctors when they inject insulin at the right time.....	136	229	54	18	8	51
AI 24.	Medicine for diabetics costs too much.....	118	136	81	112	19	30
AI 25.	It's easy to figure out how to change an insulin dose without asking a doctor about if first..	27	102	69	135	113	50
AI 26.	Diabetics will get the right kinds of help in emergencies if they have diabetic identification with them at all times.....	206	193	48	27	3	19
AI 27.	It's hard to stay on a special diabetic diet all the time.....	148	233	7	69	16	23
AI 28.	Insulin users miss out on a lot of things because they have to take their insulin.....	24	53	59	177	133	50
AI 29.	It's easy to remember to carry something with sugar in it to eat in emergencies.....	108	222	37	83	12	34
AI 30.	Following a special diabetic diet all the time keeps you from eating the way you'd like to.....	108	227	33	87	23	18

Table 21 continued

	Lot of Good	Some Good	Makes No Difference	Some Harm	Lot of Harm	Not Answered
AD 13 In your opinion how much good or harm does it do for you to test your urine the way you were told to?.....	329	78	35	2	0	52
AE 7 In your opinion, how much good or harm does it do if you do what you were told to do to prevent or take care of insulin reactions.....	332	23	14	4	1	122
AE 11 In your opinion, how much good or harm does it do for you to take your insulin as you were told to take it?.....	305	31	12	6	26	116
AG 5 In your opinion how much good or harm does it do for you to eat the way you were told to.....	368	46	17	9	9	47

	Like A Lot	Like Somewhat	Don't Care	Dislike Somewhat	Dislike A Lot	Not Answered
AD 14 How much do you like or dislike testing your urine the way you were told to?.....	126	75	117	71	58	49
AE 8 How much do you like or dislike doing what you were told to do to prevent or take care of insulin reactions?.....	220	71	52	24	8	121
AE 12 How much do you like or dislike taking your insulin in the way you were told to take it?.	150	57	83	45	43	118
AG 6 How much do you like or dislike having to eat the way you were told to?.....	100	110	55	133	54	44

	A Lot Better	Somewhat Better	Makes No Difference	Somewhat Worse	A Lot Worse	Not Answered
AD 15 All things considered, how much better or worse do you feel your life is if you test your urine the way you were told to?.....	245	99	87	11	3	51
AE 9 All things considered, how much better or worse is your life if you do what you were told to do to prevent or take care of insulin reactions?.....	312	39	19	4	1	121
AE 13 All things considered, how much better or worse is your life if you take your insulin the way you were told to take it?.....	292	48	28	5	9	114
AG 7 All things considered, how much better or worse is your life if you eat the way you were told to?.....	336	67	27	16	3	47

SCALE ANALYSIS

Twenty-three attitude items in Section I of the first questionnaire formed nine scales (at least two items each) corresponding to the specific content areas rated important by the experts. These scales were tested for conformity to convergent and discriminant correlations criteria using a sample of 304 respondents who answered at least 75 percent of the items. Missing response values were estimated using the patient's mean score for answered items in the hypothesized scale.

Results indicate that these scales do not have desirable measurement properties. The majority of the items correlate less highly with their hypothesized scale score than with some other scale score. The correlations are shown in Table 22 and summarized in Table 23. Internal consistency estimates for three of the scales were over .40 which prompted us to do some exploratory factor analyses to seek out alternative ways of organizing the items into scales.

A principal components analysis of correlations among the 30 attitude items in Section I of the questionnaire was performed. Important factors were rotated to an orthogonal simple structure. Despite rotation, the factors could not be interpreted easily. With one minor exception, each factor contained different concepts and also different evaluation dimensions (good/bad, important/not important, easy/hard, etc.). Furthermore, we could not rule out the hypothesis that much of the structure was determined by the direction of the question wording: one factor contained mostly items that required an "agree" response for a positive attitude. Our conclusion from the exploratory factor analysis is only that we did not test enough questions to distinguish among content, evaluation dimension, and method effect sources of variance.

Table 22

CORRELATIONS BETWEEN ATTITUDE ITEMS AND NINE SCALES
HYPOTHESIZED FROM DELPHI ITEM GROUPINGS

Groupings/Items	36A	61A	53A	59A	03A	99A	33A	47A	54A
36A Eats for Hypoglycemia									
A18	.28[a]	.18	.20	.12	.12	.21	.19	.12	.19
A29	.28[a]	.25	.23	.15	.17	.14	.15	.12	.29
61A Carries Diabetic ID									
A3	.13	.02[a]	.13	.30	.05	.30	.30	.14	.19
A26	.29	.02[a]	.18	.16	.18	.13	-.02	.16	.30
53A Reports Symptoms to MD									
A4	.30	.11	-.04[a]	.21	.16	-.01	.05	.13	.06
A9	.07	.18	-.04[a]	.36	.21	.27	.18	.12	.26
59A Reports Severe Insulin Reactions									
A2	.14	.18	.25	.04[a]	.12	-.05	-.05	.06	.21
A17	.11	.29	.33	.04[a]	.18	.26	.20	.14	.38
03A Fills RX-Pills									
A5	.17	-.04	.15	.14	-.20[a]	-.01	.08	-.08	.14
A22	.11	.21	.20	.20	.24[a]	.25	.07	.13	.26
A24	.06	.05	.11	.03	.03[a]	.14	.13	.40	.00
99A Takes Insulin at Correct Times									
A12	.13	.25	.20	.13	.17	.38[a]	.31	.19	.24
A14	.14	.23	.10	.13	.14	.44[a]	.22	.09	.19
A23	.18	.15	.06	.09	.09	.11[a]	.08	.16	.19
A28	.12	.16	.10	.05	.18	.28[a]	.34	.10	.04
33A Eats Meals at Correct Times									
A6	.11	.27	.13	.08	.11	.34	.31[a]	.12	.06
A13	.20	.11	.10	.00	.18	.29	.06[a]	.13	-.04
A27	.17	.06	.16	.12	.07	.19	.27[a]	.02	.15
A30	.06	.11	.02	.07	.05	.13	.36[a]	.07	.04
47A Visits Regular MD									
A7	.16	.18	.16	.10	.06	.12	-.05	.07[a]	.10
A20	.09	.13	.12	.11	.30	.18	.20	.07[a]	.05
54A Inspects Feet Daily									
A8	.25	.22	.20	.30	.10	.15	-.02	.16	.16[a]
A16	.23	.27	.16	.35	.22	.23	.12	.02	.16[a]
Not Assigned									
A1	.02	.12	.18	.16	.04	.15	.20	.06	.15
A10	.04	.29	.15	.18	.11	.35	.12	.13	.22
A11	.24	.06	.12	.08	.07	.17	.09	-.03	.18
A15	.26	.16	.15	.14	.03	.13	.09	.11	.10
A19	.15	.23	.18	.17	.08	.21	.06	.28	.13
A21	.08	-.05	.01	.05	-.01	.09	.19	.01	.00
A25	-.04	-.21	-.11	-.11	-.18	-.22	-.04	-.11	-.27

[a]Hypothesized scales and correlations corrected for overlap.

Table 23

SUMMARY OF SCALING RESULTS AND INTERNAL CONSISTENCY
RELIABILITY COEFFICIENTS FOR NINE DELPHI ATTITUDE SCALES

	Delphi Scales	Number Items	Success[a] Ratio	Errors		Reliability[b]
				Probable	Definite	
36A	Eats for Hypoglycemia	2	13/16	3	0	.44
61A	Carries Diabetic ID	2	0/16	2	14	.04
53A	Reports Symptoms to M.D.	2	0/16	3	13	.08
59A	Reports Severe Insulin Reaction	2	0/16	3	13	.08
3A	Fills RX-pills	3	3/24	9	12	.02
99A	Takes Insulin at Correct Times	4	23/32	7	2	.51
33A	Eats Meals at Correct Times	4	23/32	5	4	.44
47A	Visits Regular M.D.	2	1/16	9	6	.12
54A	Inspects Feet Daily	2	7/16	5	4	.23
	TOTALS	23	70/184	46	68	

[a]Number of definite scaling successes over the number of tests.
[b]Internal consistency reliability estimate.

ISOLATING POTENTIAL SOURCES OF ANSWER VARIATION

An important uncertainty remains about the value words used in the attitude questions (evaluation dimensions). Can any value word be used when asking about an attitude toward a particular self-care activity? Are patient attitudes organized wholly or partly around the different evaluation dimensions? If different evaluation dimensions have the same meaning to respondents, they do not need systematic attention in constructing questions for further research. If they have different meanings, the question construction and analysis strategy needs to account for them.

To address this issue, a subset of attitude questions were identified in which activity, evaluation, and direction of wording effects could be separated. The item intercorrelations were examined to determine if attitude responses were homogeneous within activity or evaluation dimensions or both. The subset of questions asked about three kinds of activity (nutrition, medication, safety) and two kinds of evaluation: help-harm (importance in achieving good health outcomes) and easy-difficult (to carry out).

Table 24 contains intercorrelations of items grouped several ways.

Entries on the Main Diagonal are correlations of items asking about the Same Activity and Same Evaluation Concept (SA/SE). For example, the entry in column 1, row 1 (.15) is the correlation of the two help/harm nutrition items.

Entries in the Triangles are correlations of items asking about the Same Evaluation Concept and different activities (DA/SE). For example, the entry in column 2, row 1 (.12) is the average correlation of help/harm nutrition items with help/harm medication items.

Entries within the Upper Right Diagnonal are correlations of items about the Same Activity and different evaluation concepts (SA/DE). For example, the entry in column 4, row 1 (.04) is the average correlation of help/harm nutrition items with easy/difficult nutrition items.

Remaining entries are correlations between Different Activity/ Different Evaluation Concept items (DA/DE).

Each reported coefficient is an average of at least two correlations summed across an approximately balanced set of positively and negatively worded items.

The general pattern of coefficient sizes is (from highest to lowest):

(SA/SE) > (DA/SE) > (SA/DE) > (DA/DE).

This suggests that both activity and evaluation concepts are determiners of attitude answer variance and that the evaluation dimensions are at least as important as the activity dimensions. Both should be considered in further measurement research.

Table 24

AVERAGE ITEM INTERCORRELATIONS CONTROLLING TYPE OF ACTIVITY AND TYPE OF EVALUATION DIMENSION

	HELP/HARM			EASY/DIFFICULT		
	Nutrition	Medication	Safety	Nutrition	Medication	Safety
HELP/HARM						
Nutrition	.15	.12	.07	.04	.10	.07
Medication		.20	.07	-.05	.01	.00
Safety			.09	-.01	.07	.10
EASY/DIFFICULT						
Nutrition				.27	.09	.08
Medication					.15	.18
Safety						.15

V. RECOMMENDATIONS FOR FURTHER RESEARCH

Three kinds of additional research activities are recommended to meet short-run measurement development goals: Final work on scale construction, identification and removal of the major forms of bias, and theory elaboration.

The present project focused on index construction and evaluation so less research is needed in this area than in the other two. Two categories of additional scale construction research are (a) finishing the development and evaluation of "promising" indexes and (b) developing and testing new measurement approaches for the concepts that were not measured or evaluated well in the present research. Almost half of the behavior knowledge indexes fall into the first category: The approaches tested in the present research look promising but more research is needed to achieve or confirm the desirable measurement properties. In the second category, needing concentrated additional research, are the behavior indexes about nutrition, self-initiated changes in the regimen, interpretation of urine sugar test results, and compliance with the oral anti-hyperglycemic medication (pill) regimen. The research did not produce acceptable measures of these constructs and there is a good deal of uncertainty about which new approaches will succeed. Extensive additional research that tests several different approaches may be needed. Finally, new research will be required to develop attitude scales because the present study was unable to devote sufficient resources to the task.

The present research has been somewhat arbitrary about the rules for scoring items and combining item responses into index scores. A task for a next research study is to explore a range of scoring approaches to identify those that contribute favorably to measurement properties without being cumbersome to use. Approaches to index construction not based on Z-scores, for example, would be desirable if

they did not detract from the validity and representativeness of the resulting index value.

A second kind of additional research effort addresses the potential biases of observation and nonobservation. The existence of substantial amounts of compliance reporting bias for some indexes poses an important threat to interpreting research results based on them. Differences between scores from two facilities, for example, could represent true trait differences or only differences in compliance reporting bias tendencies. Priority should be given to identifying test conditions that minimize the bias. Failing that, robust statistical procedures that identify and control for the effects of the bias on inferences should be developed. One bias-reducing technique was tried in the present research that did not work. There are other techniques to try that have worked in different contexts but it would be unwise to suggest that a simple solution can be developed quickly.

Additional research should replicate the present study (including the measurement evaluations) on a sample that has at least an 80 percent questionnaire response rate and a 90 percent item response rate.* Low questionnaire response rates and the existence of question nonresponse raise concerns about the generalizability of the present research findings. Did patients who would have given biased or error-prone answers just not return the questionnaires? What information is lost by not getting answers to all questions from patients who send in the questionnaire? To make sure that the indexes developed currently and in the next research are not subject to nonresponse biases, the present research approach should be repeated and efforts made to achieve high form and item response rates.

Large sample future research is needed to determine reliability and validity of responses from special target groups. These might include people with low levels of formal education, those recently completing a patient education course, selected ethnic groups, children, etc. It is not clear how to identify which target groups to study further; it does make sense, however, to identify the minimum

*These rates represent those assumed necessary for program evaluation research in general.

levels of education, English language fluency, and medical sophistication needed to give answers on the self-administered questionnaire. To the extent possible, any cultural biases should be identified and corrected.

A third and possibly most important area for future work is in theory development. The implicit theory that links educational interventions, attitudes, knowledge, compliance behavior, and the medical outcomes is not spelled out in a formal sense, tends to be static rather than dynamic, and may be ignoring important classes of variables. Yet the measures developed here (and in the future), the educational interventions to be tried, and the regimens prescribed by physicians are highly dependent on the implicit compliance theory.

Effort to state and empirically support a formal theory at this time has several likely advantages for measurement development. It can aid in establishing the predictive and construct validity of the indexes by indicating the predictions that a good measure should make and the variables that should not correlate with the measure. Careful theory specification may lead to discovery of important variables that should be measured but have been overlooked. Specification will force our attention to the kinds of <u>patterns</u> of compliance <u>over time</u> that are necessary to good medical outcomes and, consequently, what aspects of the compliance pattern need to be represented in the behavioral measures. Theory specification should also help determine the kinds of education interventions most likely to have large effects and, hence, those that might be tried in conjunction with additional measurement development research. A careful analysis of the existing data to establish the links between attitude, knowledge, skill, and behavior should aid the theory specification process a great deal.