Report of the Field Test to Evaluate a Decision Support Tool for Psychiatric Assessments in Emergency Rooms

CTG.OMH-005

CTG Working Memo Series

Center for Technology in Government
University at Albany / SUNY

I. Introduction

Each year more than 135,000 people receive emergency psychiatric care in New York State. More than half are referred to emergency services for potentially dangerous behavior. Each case requires an assessment, often made under difficult conditions, and a subsequent decision to admit the person to inpatient care or release to the community. Inconsistency in emergency psychiatric assessments have been a source of concern to the New York State Office of Mental Health (OMH) and to patients and their families. In response to this concern, a project was conducted during 1994-95 at the Center for Technology in Government (CTG). The project sought to design and test a decision support tool to assist with psychiatric assessments in general hospital emergency rooms. The tool was not intended to replace the judgment of clinicians, but was designed to prompt him or her to collect and consider all information relevant to the admission decision. The ultimate goal of OMH is to improve emergency psychiatric services by reducing the number of inappropriate admissions, and by avoiding inappropriate releases which can result in violent episodes in the community. This project represented one of several approaches OMH is investigating to achieve that goal. Work on this project began in Spring 1994 and involved staff from the OMH Bureau of Evaluation and Services Research, faculty experts from the University at Albany's Center for Policy Research, and CTG professional staff, faculty fellows, and graduate students.

Project Objectives

The project had three primary objectives:

1. **Apply methods derived from expert judgment research and group decision support technology to develop a decision model for psychiatric admissions decisions in general hospital emergency rooms.**
   The first objective of the CTG project was the development of a decision model for use in psychiatric assessments in emergency rooms. An effective and workable model requires not only careful study of relevant empirical results and existing guidance, but also requires a consensus among experts. Arriving at such a consensus is not always easy, but successful procedures (called "group decision support systems") for facilitating consensus have been developed. The Center for Policy Research at the University at Albany applied its extensive facilitation experience to this objective. A panel of ten to fifteen experts was brought together three times (in May, July, and September 1994) to define and reach consensus on the decision model. The model is based on categories of information deemed necessary to conduct a sound emergency room assessment. These categories include:

   - Danger to Self
   - Danger to Others
   - Mental Health Status
   - Functional Impairment

- Substance Abuse
- Environmental Stressors
- Potential to Benefit from Treatment
- Client/Family Preferences
- Availability of Outpatient Services

The questions in the Mental Health Status category are drawn verbatim from the Brief Psychiatric Rating Scale (BPRS). The questions in all other categories were devised by the expert panelists and the team of researchers from the Center for Policy Research. There are a total of 73 questions included in the prototype, although not all of them must be answered for each patient. There are, however, 44 "Must Ask" questions which are always required. The decision model then combines the answers to these questions into numeric scales and summaries to aid in the evaluation. The system will not present an evaluation result unless all of these questions have been answered. If a user requests an evaluation before answering all the "Must Ask" questions, the system will present those unanswered questions with a reminder that they must be completed before an evaluation result can be prepared and presented.

A select group of three experts met a fourth time, in April 1995, to evaluate the readiness of the prototype for a field test.

2. **Translate the decision model into a prototype software application that can be used by ER personnel.**
   Three versions of a computer program based on the decision model were developed by CTG technical staff. The first version was presented at the September 1994 meeting of the expert panel. A second, refined version was presented to three of the experts in a specially designed evaluation meeting in April 1995. The third version, refined following the evaluation meeting, was field tested in July 1995. In essence, the prototype, which runs on a notebook computer, prompts practitioners to enter specific information about a patient. The program poses additional questions if the information provided is not deemed adequate for a sound assessment using the underlying decision model. The computer program then integrates the information and presents both summary and detailed reports in the form of statistical profiles and narrative descriptors. A more complete description of the prototype may be found in Appendix A.

3. **Field test the prototype to assess user reaction and usability as a support tool in ER practice.**
   The field test took place at Westchester Medical Center during July 1995. An Institutional Review Board gave approval for the research and the participating physicians and patients all signed consent forms. The field evaluation comprised a first assessment of the structure, value, and usability of the tool in a hospital setting. The field test included an examination of the possibilities for three different uses of the tool: for professional training, as an interview aid, and as a

post-interview recording/reference device. Additional testing of the prototype in other hospitals and a full evaluation of its validity, impact on the quality of assessments, and impact on subsequent admission/discharge decisions is planned by OMH, but was outside the scope of the CTG project.

II. The Design of the Field Test

The field test was designed as a first look at an automated decision support tool in a real life setting. It was designed to gather reactions from practicing clinicians that would lead to refinements and further testing. With this specific purpose, the test was quite limited in both scope and duration. It was not intended to be a comprehensive assessment of the prototype in operation, nor to be a study of the quality of decisions made.

The field test took place during a six-day period in July 1995 at the Psychiatric Institute of Westchester Medical Center in Valhalla, New York. Westchester Medical Center is not typical of psychiatric ER services around New York State. It is a large and multi-faceted teaching hospital with a sophisticated, well-staffed psychiatric emergency room within the Psychiatric Institute. The Psychiatric Institute also includes a mobile crisis team which conducts crisis/triage in the community using a three-item crisis/triage rating scale created by the clinical staff and in use since 1982. Each of the three items on the scale (dangerousness, support system, and ability to cooperate) comprises five descriptive conditions which anchor associated numerical ratings. Westchester was chosen as the field test site because its expertise and comprehensiveness would ensure a very rigorous evaluation. The field testers included three physicians on the staff of the Institute including the Director of the Comprehensive Psychiatric Emergency Program (a psychiatrist) , an experienced attending psychiatrist, and a second year resident doing his emergency program rotation. The Director was also a member of the expert panel which had helped define the underlying decision model. All three were experienced computer users.

The prototype was delivered to the test site already installed on a high-quality notebook computer, which could be used easily in any room by any trained member of the staff. The testers were trained in the use of the system by the director of the Center for Policy Research and a representative from the OMH Bureau of Evaluation and Services Research who had participated in the entire project and was familiar with the underlying issues, and with the Westchester site and staff.

The testers opted to use the system after conducting their regular assessment interviews in which they continued to use their existing forms and procedures. The clinicians then entered data about the patients into the prototype and compared their own assessments to the system's evaluation of the data they had entered. During the period when the prototype was in place, seven sample cases were collected. Five cases were patients who were evaluated in the emergency room during the test period. Two

additional cases were entered into the system from the records of patients who had been seen just before the test began. For the first three cases, the OMH evaluator and the physicians used the prototype together, discussing the data collected in the interviews and matching it to the questions contained in the prototype. After the first three days, the system was loaded on the desktop computer in the Director's office and additional cases were added until the end of the test period.

Field test results were collected through semi-structured interviews conducted by two members of the CTG research staff. The interviews were conducted the day after the end of the test period and included all three practitioners who had participated in the test. The interviews covered the following topics:

General reaction to the technology:

- Ease of Learning
- Ease of Navigation
- Ease of Data Entry
- Clarity and layout of screens
- Bugs

Effectiveness of the prototype for decision support:

- Assessment of logical structure (e.g., the organization and flow of questions)
- Time needed to complete an automated assessment
- Appropriateness of the questions
- Appropriateness and clarity of language
- Usefulness of outputs
- Completeness of coverage (Does it ask all the pertinent questions?)
- Potential for integration into existing ER practices
- Value-added characteristics
- Negative characteristics

Overall assessment of potential for future use:

- Potential users
- Potential uses
- Needed changes
- Recommendations for enhancements and new features
- Recommendations for further testing

III. Field Test Results

The physicians who participated in the field test made many observations and recommendations based on their use of the prototype. Their comments are summarized below:

Advantages of the system as currently designed:

- For the sample patients, the system-generated final evaluations agreed with the clinician's professional opinions.
- The final output added value to the ER assessment by encouraging the practitioner to reflect on the factors which contribute to the admission decision. The comparison of the computed scale scores and the single global measure in each category is especially useful since it can reveal disparities and gaps in the data. Such a disparity should prompt the clinician to explore the reasons for this difference and to acquire additional information or seek consultation.
- For the inexperienced practitioner, the system acts as a comprehensive "checklist" of information which is important to collect and include in a psychiatric evaluation.

Disadvantages of the system as currently designed:

- The system is a cross between a crisis decision support and an aid to a full psychiatric evaluation, but does not fully meet the needs of either use.
- As a crisis decision tool, it is too long and complex to be practical in the ER or in the field. It could be improved by paring down the questions to a smaller set of "must ask" questions which focus on an assessment of dangerousness to self or others.
- As a full evaluation aid, the prototype lacks internal logic which would lead a clinician through an evaluation that is meaningful for a particular patient. In other words, the system neither "branches" to new questions based on the answers to previous questions nor leads to conclusions such as tentative diagnoses. Instead, it requires simply that certain questions always be answered.
- The prototype uses a sophisticated psychiatric vocabulary implying that it is designed for professionals with a high level of psychiatric training. However, the tool is also highly structured and has often been described as one designed for less expert professionals. This discrepancy makes it difficult to evaluate how useful it would be for various kinds of users. A novice user might need to follow a more structured, lengthy set of questions, but those questions must be translated into more common language. An expert could deal with the expert vocabulary, but would not need the high degree of structure in how the questions are presented and answered.
- The questions lack descriptive "anchors" which would help assure the reliability of the assessments. General terms like "mild," "severe," and "moderate" used by the prototype to anchor points on the scales are subject to varying interpretation.
- The system collects demographic characteristics of the patient but does not use them in presenting the evaluation results. This is very important information which should figure in the final decision about a case.
- In the evaluation screens, all categories of information are displayed graphically on the same zero-to-six point scale. The testers felt this could be improved by weighting certain categories (e.g. dangerousness to self or others) to show their relative importance.

Reactions to the technical implementation:

- The system is easy to use. The Windows interface is readily recognizable and makes the system immediately usable with little training. However, there were some technical problems. The minimize and maximize buttons do not act in a Windows-like way and there is no intuitively obvious way to return to the first screen of data (demographics) for a case-in-progress.
- A listing of existing records, organized by patient name or some other identifier, should be available so that the records of specific patients can be retrieved and reused easily. This database of patient records must have both read and write protection to assure confidentiality.
- The "Must Ask" button should be highlighted more vividly. Except for the red color, it looks like all the other buttons. Similarly, in the evaluation screen, the "must ask" questions in the dialog box should be displayed in bold or in color.
- In the evaluation portion of the application, a bar graph of the computed scale scores should be presented first, followed by a second multiple bar graph which compares the scale scores to the global scores.
- The underlying scales assign scores to some questions which are less finely differentiated that the user responses. For example, a user-assigned value of 2 means something different (to the same user) than a user-assigned value of 3, but on some questions the formula which computes the composite scale score gives them the same value.
- Scores and summary evaluation questions and answers can be erased and edited by the user on the screen. This should not be allowed on the evaluation screens which simply report results that are calculated from the values assigned to specific questions.

IV. Recommendations

Based on this early test of the prototype, we recommend that OMH consider the following next steps.

1. Test the prototype in several additional settings before making major modifications. Westchester Medical Center was an excellent site for the purposes of this particular test, because it is unusually rich in resources and expertise. It is now important to test the prototype in a variety of other settings where emergency psychiatric decisions are made. Two settings are especially recommended: a large general hospital with psychiatrists on staff but without a psychiatric ER, and a rural hospital without staff psychiatrists. We recommend following the same general evaluation protocol as used in this test and comparing the results across sites. Once this larger body of test data is collected and analyzed, the prototype can be revised more effectively.

2. After additional testing, revise the decision model and the user interface for a clearly defined intended user and setting. The intended system user needs to be more clearly defined and reflected in the user interface. No single system can be effective for every potential user, use, or setting. A psychiatric professional, for

example, would need less prompting and a different vocabulary than would a non-psychiatric medical professional. A novice or a student would demand yet a different set of system characteristics.

The physicians in Westchester said that the system, as is, cannot be used effectively without psychiatric training at least at the level of a psychiatric nurse. The Mental Health Status questions especially (which are taken verbatim from the American Psychiatric Association's BPRS) require a vocabulary and level of sophistication not likely to be found in a general ER. The BPRS vocabulary is especially problematic since common words (like "anxiety") have special meaning within psychiatric practice. The testers told us that doctors and nurses in a general hospital ER are not likely to recognize a psychiatric level of anxiety from the ordinary level that any person would have about being in an emergency room.

The testers stated that BPRS questions are useful for research but not as an aid to clinical practice. They recommended that these items be replaced with a few concrete behavioral descriptors regarding the patient's use of language, facial expressions, and so on. The general hospital practitioner's observations of these specific behaviors are more in keeping with their training and experience and would give good indicators of mental status which a psychiatrist could evaluate later.

The presumed setting in which the tool would be used would also make difference in how it is redesigned. The Westchester physicians felt that the system is not a useful addition to a sophisticated psychiatric ER. In such a setting it is too long for a screening tool and not detailed enough for a full assessment tool. In Westchester, at least, it duplicates data already collected on paper during an unstructured assessment interview and adds work, but little additional value, when completed as a separate additional step. However, in a smaller hospital, or a rural setting, or even a non-medical setting (like a social services office) where there is less access to psychiatric experts, the physicians thought the tool might prove to be very valuable.

3. After additional testing, revise the system to achieve a specific purpose. It is possible that several purposes could be served by different versions. With appropriate modifications, the field testers recommended the following uses for the system be considered:

- *A training tool for psychiatric residents, psychology students and interns, medical students, nursing students, social work students, emergency medical technicians, social services office staff, and other human services and mental health professionals.* The prototype as it is currently devised is much more than a crisis-oriented decision tool. It contains most elements of a full psychiatric evaluation. As such, it has very good potential to become an automated training device for professionals in a variety of fields who must prepare to work with people who have mental illnesses. It could be used to train practitioners to conduct a full psychiatric assessment, an assessment of dangerousness, or other kinds of decisions. The

calculation of scale scores which can be compared to the student's global assessments offers opportunity to explore the relative importance of various categories of information and the way in which the student interprets patient or client data. The comprehensive set of categories and questions makes it possible to present students with a variety of "test" cases for evaluation and discussion. Moreover, once information is captured in electronic form, it can be compared, analyzed, and integrated with other information. All of these functions could be useful in an educational application.

- *A crisis decision-support tool for non-psychiatric professionals such as general hospital emergency room personnel.* According to the Westchester physicians, the system could be streamlined and focused on the triage or immediate admissions decisions. It needs to be "doable" in minutes. For this purpose, it need do no more than reliably answer the question, "Should this person be held in the hospital for further psychiatric evaluation, observation, or treatment?" The testers at Westchester made specific recommendations for the information which should be collected by such a tool. Their recommendations are presented in Appendix B. They thought that such a system might best be used after, not during, a more traditional interview, but recommended that various approaches be tested in further field trials.

- *A tool to directly aid or structure a full assessment interview*. This is a possible use for the non-psychiatric professional. However, the testers expressed strong reservations about this application since it introduces a high degree of structure into the interview. They believed it would take too much time and be too distracting to both practitioner and patient. They thought it would be very difficult to try to follow a naturalistic, patient-led, unstructured interview (which these physicians prefer) and still find and answer the questions in the system. Nevertheless, for non-experts, they thought this use of the system could be valuable either in the interview itself or as an immediate behind-the scenes cross check that all critical information has been collected.

4. Revise the list of "Must Ask" questions to match a specific purpose. The "Must Ask" questions need to better match a specific user and purpose. For example, if the tool is to be used for a triage/crisis decision, fewer questions need be asked. If the tool is to be used for training or support for a full psychiatric assessment, then more topics need to be included; history of the current problem, specific stressors, diagnosis, and interpersonal problems are possible innclusions. In this case, the BPRS questions in the prototype might be replaced with a module for recording in narrative form the results of a more standard mental status exam.

5. Use weights to indicate the relative importance of different information. The expert panel recognized that the relative importance of categories remained an outstanding issue. The field test results suggest some ways to apply weights to make the presentation of results more intuitive. To illustrate, consider the bar charts which

summarize the evaluation of each patient. The bar charts presented in the evaluation screens show the score for each category (danger to self, danger to others, etc.) on the same six-point scale. Since no scale is given more weight than any other, this "washes out" the relative importance of the different measures. A "3" on dangerousness to self or others may be far more important to the ER admissions decision than is a "3" on functional impairment. Yet both categories will be presented in an identical way in the evaluation screens. The field testers recommended that some scales be weighted for importance and thought that pre-defined demographic profiles should also carry higher weights in the final scores.

6. Disasggregate data where the interaction among data categories can present a false composite picture. The system currently presents the functional impairment scale score as a composite of the individual items pertaining to both functional impairment and social support system. While it is true that social supports often mitigate functional impairment, the clinicians believed this aspect of the system resulted in a false impression of functional impairment *per se* and preferred to look at the scores separately. Similarly, environmental stressors should be scored and reported separately.

7. Consider enhancements that take advantage of other electronic information resources. The fact that this prototype uses the power of information technology, means that links to other information resources are possible. The testers recommended three enhancements that take advantage of this feature:

- Link the system to other hospital information such as previous admissions records.
- Add a notepad to the system to allow the clinician to add narrative information to supplement the numeric scores.
- Link to commercial databases of related information. For example, some prescription drugs cause psychiatric symptoms. It would be very useful, especially in the general hospital setting, to be able to screen the patient's medications against such a database.

Appendix A
Brief Description of the Prototype

The prototype system is a stand-alone Microsoft Windows-based program, written in MS Visual Basic that generates statistical profiles for a client in each of nine categories:

Danger to Self
Danger to Others
Mental Health Status
Functional Impairment
Substance Abuse
Environmental Factors
Potential to Benefit from Treatment
Client/Family Preferences
Availability of Outpatient Services

These profiles are indicators of the risk in each of these categories, and are generated from users' answers to a set of approximately seventy-five questions about the client. Not all questions need be answered about a client; the system is capable of handling situations in which only partial information is available.

The prototype system has two components: a user interface and an evaluation report. The user interacts with the prototype system using a Microsoft Windows-based interface.

After entering the client's ID number and the user's initials, the program prompts the user for demographic information about the client (Figure A-1). The clinician can enter this data at this point if it is known, or it may be entered at a later time.

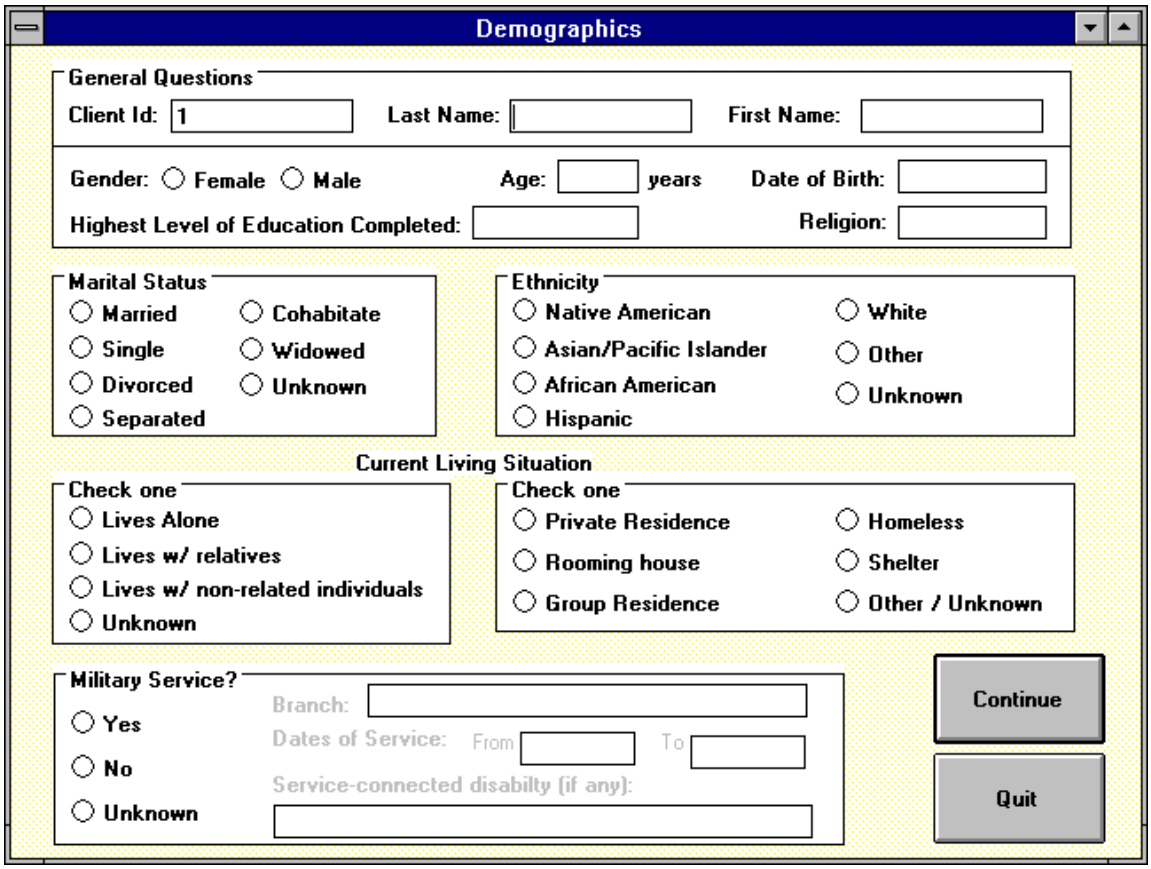

Figure A-1
Client Demographic Characteristics Data Entry Screen

The user then answers questions about the client (Figure A-2). Each question in the system is identified with one of ten categories. Questions in a given category can be displayed by pressing the appropriate button.

The user can then answer questions about the client in any order, skipping between categories as desired. Each question is answered either as "Unknown/Not Applicable", or on a severity scale of 0 through 6. Indicators of the severity of each of the numeric scales are indicated below the sliding scale.

Certain questions (such as whether the client has ready access to a weapon) contribute to several of these profiles. In "Free Entry," only the questions that are classified in the category are displayed in that category; in "Directed Entry", all the questions that contribute to the statistical profile of the category are displayed.

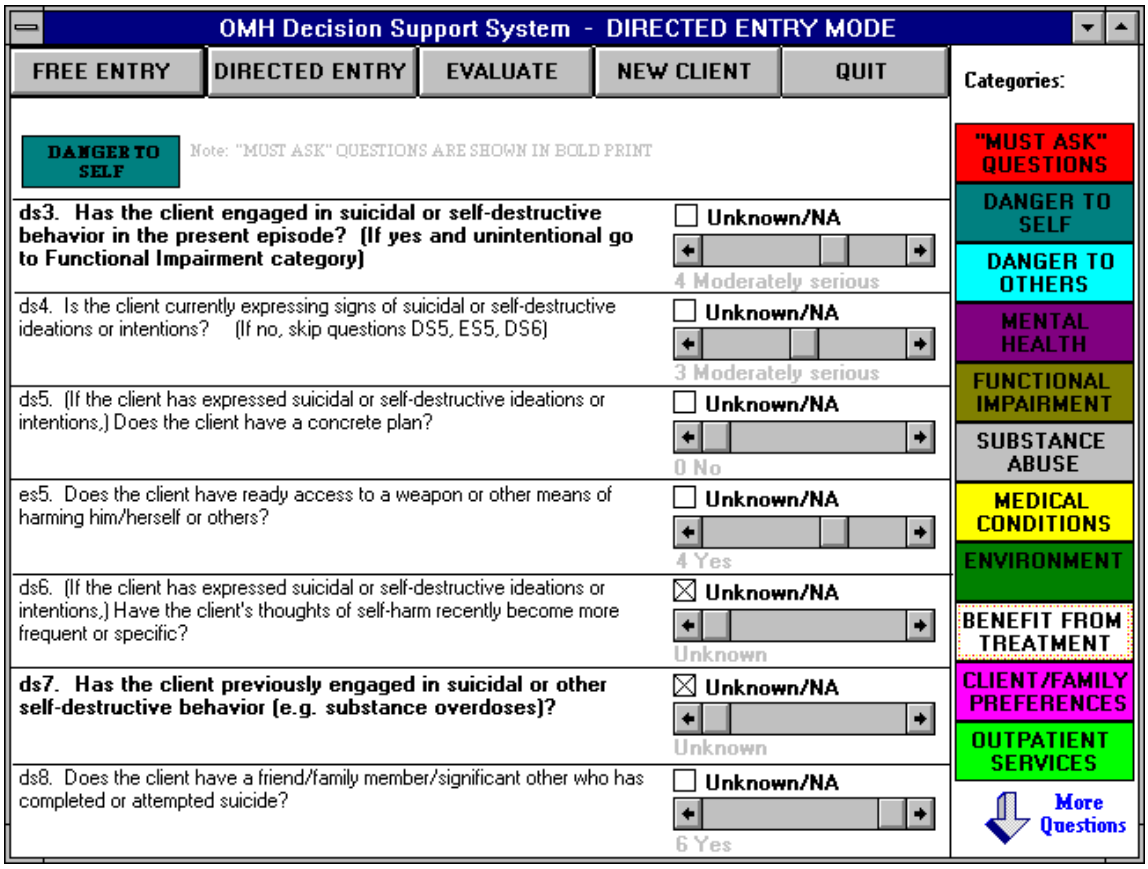

Figure A-2
Data Entry Screen, Partial Data

When the user has entered values for the questions that he or she knows about the client, the user may then request that the system evaluate the client. At this point, the system might not have enough information to create a meaningful profile. If this is the case, the system will require that the user enter answers (even if they are "Unknown/Not Applicable" answers) for certain important questions.

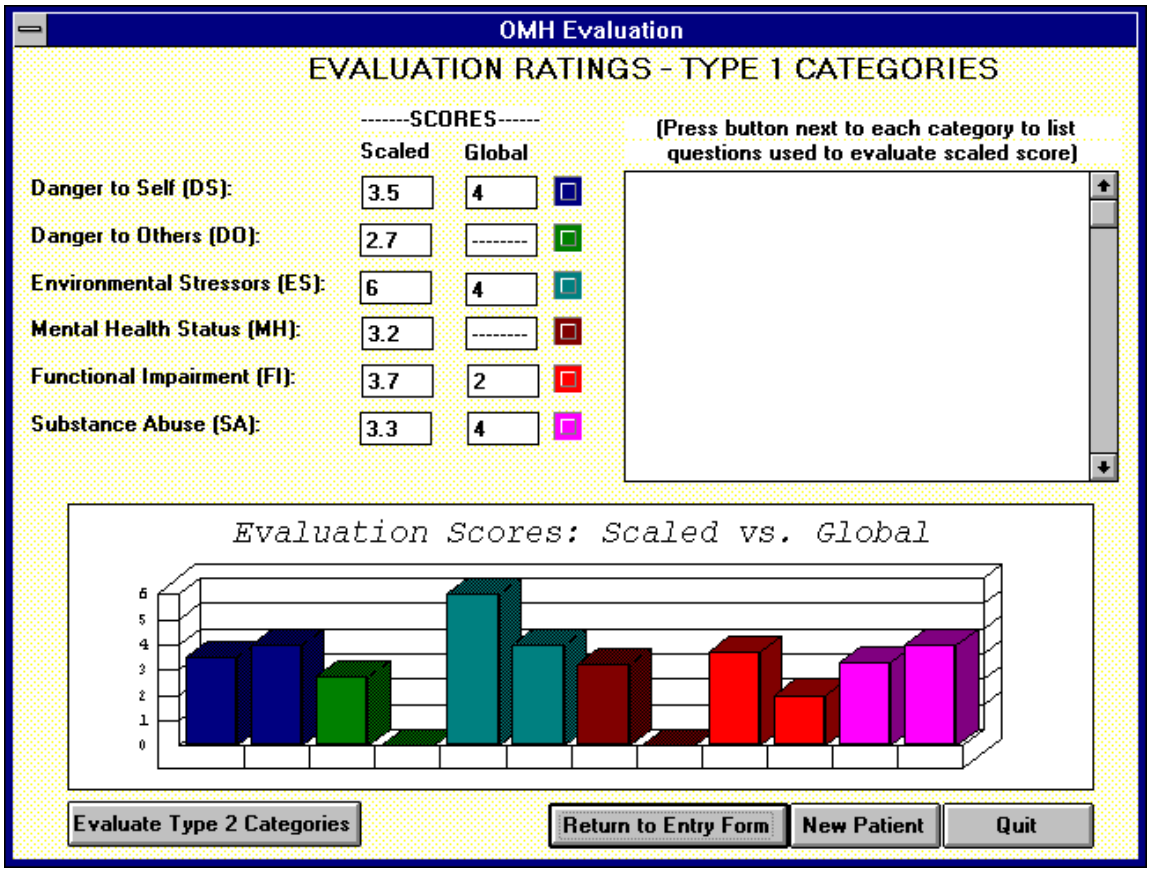

Figure A-3
Evaluation Report Screen

Scores are reported on a numeric 0-6 scale and as a bar chart (Figure A-3). For certain categories, two scores are reported: the scaled score is a weighted value determined by a number of questions; the global score is the value of the clinician's answer to a "summary" question pertaining to the category.

Appendix B
Possible Content of a Streamlined Crisis Decision Support Tool

Based on their use of the prototype, the physicians who participated in the Westchester field test identified questions that could be included in a streamlined tool for emergency psychiatric assessments. This is a very preliminary list, presented to illustrate one way that the tested prototype might be revised to become a more practical tool.

Retain these existing questions (with modifications in *italics*):

do	3	Has the client engaged in any aggressive behavior (including homicidal and assaultive behavior, or domestic violence) in the present episode?
do	4	Is the client currently expressing signs of homicidal, assaultive, or violent ideations or intentions? (If no, skip questions DO5,ES5, DO6)
do	5	(If the client has expressed homicidal, assaultive or violent ideations or intentions,) Does the client have a concrete plan?
do	7	Has the client previously engaged in aggressive behavior (including homicidal and assaultive behavior or domestic violence)?
ds	3	Has the client engaged in suicidal or self-destructive behavior in the present episode? (If yes and unintentional go to Functional Impairment category)
ds	4	Is the client currently expressing signs of suicidal or self-destructive ideations or intentions? (If no, skip questions DS5, ES5, DS6)
ds	5	(If the client has expressed suicidal or self-destructive ideations or intentions,) Does the client have a concrete plan?
ds	7	Has the client previously engaged in suicidal or other self-destructive behavior (e.g. substance overdoses)?
ds	8	Does the client have a friend/family member/significant other who has completed or attempted suicide?
es	3	How well can interested family, friends and others (including treatment providers) be expected to help to protect the client from harming him/herself or others? ***(Make this into a statement, not a score or a yes/no question. Ask instead such questions as: "Did family or friends accompany the patient to the ER?" "Did they stay?" "Did they ask questions?")***
es	4	What is the degree of stress in the client's family and social environment (e.g. Is there current or a history of abuse/domestic violence)?
es	5	Does the client have ready access to a weapon or other means of harming him/herself or others?
es	6	Has the client been a victim of abuse or domestic violence?
mc	1	Might the client have a medical condition that is presenting as mental illness? ***(Add a prompt asking the user to describe medical problems)***
mc	2	Does the client have acute or chronic medical conditions which require referral? ***(Ideally, linked to a reference database)***
sa	4	Is the client presently intoxicated or in the state of withdrawal? **(*Add: withdrawal from what drug and ideally link to a reference database of drug information*)**

Make all demographic items mandatory, plus:

- Add: “Era” for military service.
- Break “living with relatives” into two parts: “living w/spouse” and “living w/other relatives.”
- Add immigrant status.
- Add employment status.
- Add student status.
- Add “needs [language] interpreter.”
- Define high-risk demographic profiles and when a patient matches one of them, that fact should be reported on the final evaluation screens.

Retain the following concepts from the BPRS questions but reword them into concrete behavioral descriptors:

- mannerisms
- grave disability
- thought process
- hallucinatory behavior
- conceptual disorganization

Add a section where current medications are listed.
Ideally, link this to a database of drug information.

Appendix C
Evaluation Participants

Westchester Medical Center
Lawrence Levy, MD
Eric Grabstein, MD
PeShein Wynn, MD, MPH

Center for Policy Research
Jeryl Mumpower, PhD

Office of Mental Health
Bruce Way, MPH

Center for Technology in Government
Anthony Cresswell, PhD
Sharon Dawes, PhD

Center for Technology in Government
University at Albany / SUNY
1535 Western Avenue
Albany, NY 12203
Phone: 518-442-3892
Fax: 518-442-3886
info@ctg.albany.edu
www.ctg.albany.edu