MEASURING THE QUALITY OF CLINICAL ENCOUNTER DATA IN COMPUTER-BASED PATIENT RECORDS

by

Judith R. Logan, M.D.

A THESIS

Presented to the Division of Medical Informatics and Outcomes Research
and the Oregon Health Science University
School of Medicine
in partial fulfillment of the requirements for the degree of
Master of Science

May 1998

School of Medicine
Oregon Health Sciences University

CERTIFICATE OF APPROVAL

This is to certify that the Masters thesis of

Judith R. Logan, M.D.

has been approved

[redacted]

Professor in charge of thesis

[redacted]

Member

[redacted]

Member

[redacted]

Member

[redacted]

Chairman, Division of Medical Informatics and Outcomes Research

Associate Dean for Graduate Studies

TABLE OF CONTENTS

Acknowledgments ii

Dedication iii

Abstract iv

Measuring the Quality of Clinical Encounter Data in Computer-Based Patient Records 1

Studies on Data Quality in CPRs 5

Defining Quality 9

Measuring Quality 12

Completeness and Correctness 16

The Research Objective 21

Methodology: Overall Design 22

Methodology: Patient Encounters, the test material 23

Methodology: The Subjects 28

Methodology: The CPR and development of the encounter forms 30

Methodology: Data Elements, making the gold standard 33

Methodology: Data Errors 39

Results 42

Discussion 46

TABLE OF CONTENTS (Continued)

Conclusions 53

References 55

Appendix A: Encounter Forms 62

Appendix B: Abstracted Data 79

ACKNOWLEDGMENTS

My thanks to many people who made this project possible: to Dr. William Wojeski and Dr. Susan Price for help in developing the gold standards; to Dr. Michael Zacks for abstracting data and obtaining patient material; to Dr. Jon Blackman, Dr. Martha Gerrity, Dr. John Christensen, Dr. Barry Egener, and Dr. Jan Madill for helping obtain patient material; to the clinicians who gave up valuable time from their practices to serve as subjects; and to the members of my thesis committee for their excellent guidance and assistance.

DEDICATION

To the memory of Denis S. Avery, M.D.,

physician, scholar, loving father and husband.

May heaven be the Clackamas River with an exceptionally good run of Spring Chinook.

ABSTRACT

With the increasing computerization of patient records comes a reason to define quality in those records. This interesting task includes the evaluation of different methods of data capture, from free text to structured data entry. Quality in computer-based patient records (CPRs) may have several attributes, including completeness and correctness. This work defines a methodology for measuring these attributes. Two videotaped patient encounters were viewed by eight clinicians. Records of each encounter viewed were made either by dictation or using a CPR, making use of encounter forms which combined structured data elements and templated free text. The reports which resulted were compared with the gold standard for those reports as defined by a group of experienced clinicians, based on equivalence with the "elements" in that gold standard. Results were grouped by method of recording and showed no significant differences for completeness or correctness although there was a trend for the dictated records to be more correct. For completeness, the mean ± SE (95% confidence interval) for dictation was 67.7 ± 12.7 (55.0-80.4%), for the CPR was 69.9 ± 4.5 (59.2 - 80.6%). For correctness, the values for dictation were 98.2 ± 1.2 (95.3 - 101.1%) and for the CPR were 92.6 ± 2.5 (86.6 - 98.5%). This method appears to meet published criteria for research on data quality in CPRs.

MEASURING THE QUALITY OF CLINICAL ENCOUNTER DATA IN COMPUTER-BASED PATIENT RECORDS

Judith R. Logan, M.D.

The current push for computer-based patient records comes for many reasons. The belief that the omnipresent spiraling costs of health care can be controlled with the aid of computers, and with the data from computerized patient records, provides significant financial support for this endeavor. This is not necessarily the most compelling reason for computer-based patient records (CPRs). Studies conducted by longitudinal observation of aggregate data (outcomes studies) hope to add to the knowledge of medicine in areas that controlled trials are unable to reach.[1, 2] Such studies are virtually impossible without computer-based systems. Reimbursement for services depends increasingly on specific documentation. Legal issues added to the inherent desire to provide better services -- to practice better medicine -- increase the need for auditable and audit-proof records.[3] Then there is the primary use of patient records: to record information about the patient and care given either as a reminder to the clinician of prior encounters or as a communication to future providers of care. Clearly there is hope that computer-based systems will improve these records, at a minimum by increasing accessibility of the charted material. The recent marked technological advances in computer hardware and software make the time right for this effort. However, there are some hurdles to be overcome, including decisions on standardized technology that allows for interactivity, reusability, and portability within and between systems. Usability issues are particularly important given constraints on clinician time in modern health care settings. Questions as to what data to record and how it should be recorded also remain to be answered.

If the gold standard for clinical data in a computer-based patient record is the state of the patient,[4] then how does the ideal record look? Is the patient state best represented by finely granular descriptions and terminologies, perhaps codified, and deposited easily into a data warehouse for subsequent analysis? Or, is it represented by inferences, summary statements and broad overviews for rapid assessment of status? We know little about the ideal record from the standpoint of communication. How should the data be structured and how is that data best captured?

Most computer-stored encounter data -- that is, the patient's history and physical, and the clinician's assessment and plan based on a particular encounter between the patient and clinician -- is in free text (natural language) statements, captured either by keyboard entry or by transcription of dictation. This allows for capture of anything the patient says or any physical finding, and of the decision-making process including the clinical impression and plan. It is rich data in the sense that it can capture subtleties and nuances in histories and findings. It is easy to collect. It is not inexpensive, however, if transcribed. If dictated, there may be a significant time delay between the dictation of the record and the appearance of the transcript of that dictation in the computerized chart. If entered by keyboard by an average clinician, it is not fast. Speech recognition technologies may supplement or replace transcription such that the expense is less severe and time less an issue, but the technologies are still emerging and not in wide use at this point. Even if they were, this free text data cannot be codified and it is not searchable or retrievable except with natural language processing tools, which may deal well with syntax but not with the context or meaning of words which are so important and complex in medicine; medical discourse does not "yield to linguistic analysis."[5]

An alternative to the capture of encounter data as totally free text is the use of templates which guide the user to enter text into a structure, and may include the use of predefined or "boiler-plate" text excerpts. This approach solves some of the contextual problems that exist with natural language processing, and may speed and standardize data entry to some degree. Experience with templates in paper records shows that the data collected may be more complete,[6] but that data still cannot easily be codified and quantified.

A third alternative for capture of data into the record is by structured data entry, where in addition to the structure of templates is added a restriction on the number of responses to any item. Choices are usually made by radio buttons, with check boxes, or chosen from single or multiple pick lists. This data is conceivably entirely codifiable. It is quantifiable. It may be finely granular or broad. It can be collected at the bedside. However, it is not rich since the choices are necessarily restricted, it may not be fast to collect, especially early in the training period, and significant training is usually required. It works best in encounters where the data to be collected is limited, quantified and definable.[7, 8]

Which method of data capture will produce the best quality computer-based patient record? The ASTM standard E-1384[9] on the content and structure of the computer-based patient record reflects general sentiment in the statement that "for some domains (for example, history and physical) there is lack of both code systems and experience with structuring. There is also yet no empirical basis for deciding how much to structure and how much to leave as free text. Given these realities, historical preferences and the mass of existing free text information, the CPR must accommodate both structured and free text reporting for the foreseeable future." While this is undoubtedly the appropriate conclusion for this time, efforts to define and measure quality may result in a less empirical basis for choice of data capture methods in the future.

STUDIES ON DATA QUALITY IN CPRs

It would be inappropriate to ignore past studies on quality in paper-based records in a review of articles on quality in computer-based systems. Many of the issues are the same: Are encounter records a reflection of the quality of care provided?[10] Can the records be used for audit purposes? Are they a reflection of the actual patient encounters? Do they reflect the process of care? How complete are the encounter records? How accurate (correct) are they? Regardless of the methodology chosen, the message from paper-based records is that these records contain significantly less information than does the actual encounter. [6, 11-14 15-17]

A notable exception to this was the study by Duggan and associates[6] using paper charting forms which provided a structure for collection of age-specific "process items" for pediatric health maintenance examinations. The recorded encounters were also observed. Not only was there less difference between recorded and observed practices when structured forms were used than when free text forms were used, -- i.e. the structured form was a better estimate of performance -- but also overall care as measured by the number of items performed improved with the structured forms. The price paid for this improvement, however, was the tendency to overdocument physical findings. The authors noted that the format of the physical examination section of the structured record made it prone to this occurrence, with the use of a check-off mechanism for documenting normal or abnormal findings.

Gerbert and Hargreaves[11] used 20 criteria felt to be important in the evaluation of COPD patients, including symptoms, signs, tests, treatments and patient education criteria, to compare methods for evaluating performance. They first videotaped patient encounters and after the encounter audited the chart and interviewed both the patient and physician. Significantly more content was recorded by interview of either the patient or physician as compared to either the chart or videotape. What, then, is the gold standard for content of the encounter? Their conclusion was that the absence of symptoms was often not documented, and even not discussed in the encounter on the assumption by the clinician that issues which are not revealed by the current complaint are negative. What is the value of what is not said?

Searching the literature for past research on the quality of CPRs was made easier with the publication of a review by Hogan and Wagner in September of 1997.[4] These authors conducted an extensive search of the literature through February 1996, looking for articles dealing with accuracy in CPRs. They developed a scoring system for critiquing the studies, then summarized the results for completeness and correctness as reported. Their search located 20 articles that reported results of 26 studies of accuracy in 19 unique CPRs and serves as a reference point for future studies.[5, 7, 8, 18-33] The majority of these articles are either studies of the errors in transcription of data from paper to computer databases,[18, 21, 22, 26, 30] studies of the quality of summary or encoded data such as problem lists, keywords, or diagnoses,[5, 19, 23, 27, 29, 31] or compare paper-based and computer-based records.[20, 24, 28]

The report from Kuhn and associates[7] compared template-driven ultrasonography reports with free text reports; they defined quality based on the completeness of a defined data set required for specific technical exams. No errors of omission were found in their structured records as compared to rates up to 20% in free text records. Gouveia-Oliveira et al[8] also evaluated technical examinations for which a defined data set exists. They found that for endoscopic examinations of specific lesions there was twice as much information in structured data entry records as in free text records, and credit this result to the "reminder" function of structured data entry charts. These studies have again established that in a setting where a well defined data set exists, templated and structured data entry records are equal to or better than free text records, at least in the measure of completeness.

Wagner & Hogan[32] studied errors in computerized medication lists. Notable in this study was the search for sources of those errors found. Correctness overall was 83% but with .37 medications missing per patient. Causes of the errors included patient-initiated changes in medications (36%), data entry errors (8%), uncaptured changes made in the paper record (13%), and unrecorded changes made by clinicians within the clinic (14%) or outside of the clinic (25%). They suggested expanding the scope of the CPR to all clinicians who can make changes in the patients' medications.

Pringle, Ward and Chilvers[33] studied four general practice groups in England known to have a high commitment to use of the EMIS computer system. Concurrent paper-based records are routinely kept. Their study had several parts, looking at different aspects of quality. These included a study of the *consistency* of data recorded on the computer, looking at the prevalence of ten diagnoses compared with national data; at *completeness* of lifestyle and socioeconomic data, looking for entries for smoking habits, alcohol consumption, occupation, social class and ethnicity; and of *accuracy* in the computerized records of the diagnoses of two diseases, diabetes mellitus and glaucoma. Results of the study of completeness showed that smoking habits were recorded 52.1% of the time, alcohol consumption was recorded 37.5% of the time, occupation was rarely recorded and no records showing ethnicity or social class were found. Accurate diagnoses of diabetes and glaucoma were present 96.7% and 92.3% of the time with no incorrect diagnoses made for these two diseases.

The fourth part of their study looked at the completeness of the records of patient encounters and consisted of two parts. Paper-based and computer-based records were compared for 1000 patient encounters, then both types of records were compared against the encounter videotapes for 200 encounters. Using the videotapes as the gold standard, both paper-based records and computer-based records failed to capture much of the information in the encounters. The computer-based records were more complete only with respect to recording of the diagnoses, while both systems recorded referrals well.

DEFINING QUALITY

Quality in medical records has been described by the Institute of Medicine as having the attributes[34] of legibility, accuracy, completeness, and meaning. There is a potential for improvement in each of these quality attributes with computer-based patient record systems. Computer-based systems can improve legibility by eliminating handwriting, but other issues of legibility still exist: in systems allowing keyboard entry of free text, unusual or inappropriate abbreviations can still be used, and the format of the records on screen or in print contributes to legibility.

Accuracy, defined here as the correctness of the data,[a] can potentially be improved by computer-based record systems by eliminating any intermediaries between the clinician and the final record, that is, by eliminating the transcriptionist for dictation or clerical personnel for entry from paper-based notes. In addition, using logical rules for entry of data into fields gives the ability to block inappropriate entries and to present alerts to the user. For instance, a field for birth date can block entries which are not numerical, are not in the correct format (xx/xx/xx or other chosen format) and which are not dates prior to the current date. If fields have well-defined choices for the values appropriate for those fields, then structuring the data entry with check boxes, radio buttons and pick lists can also improve accuracy.

Completeness in a computer-based record depends in part on agreement among users about what constitutes the core data elements for a record and accuracy depends on knowing the correct values for those elements. The use of templates in paper-based records has been shown to improve completeness[6] where well-defined data sets exist, and computer-based templates and higher structuring can be expected to do the same. Both completeness and correctness can potentially be increased by the use of editing rather than rewriting of many sections of the chart. However, the calculation

---

[a] Accuracy has alternately been defined as equivalent to completeness or as equal to the product of correctness and completeness. For clarity, the second definition will be used for the remainder of this paper.

of completeness and correctness are not simple in the context of the complexity of data found in the clinical encounters or errors found in clinical records, a topic which will be illustrated later in this paper.

Data dictionaries and coding systems using controlled vocabularies can improve the consistency of the names and descriptions found in patient records. Computer-based systems facilitate wider use of these vocabularies and dictionaries. The consequent consistency allows for improved meaning to the users of clinical data, including the communications between clinicians and in retrieval of data for studies or audits.

Ultimately, the quality of a patient record must be judged by whether or not that record serves the purpose for which it was intended, and patient records serve a number of purposes.[35] They provide documentation for billing, record information for the next encounter by the same clinician, communicate to other clinicians who provide care for the same patient, serve as a record for chart audit for quality improvement processes and legal purposes, and contain research data to feed medical science. While the gold standard may be the same for all of these purposes, that the record reflects the state of the patient, the optimal number of data elements, optimal granularity of these elements, and optimal presentation in the record may vary with the purposes.

For example, billing requires the identification of defined classes of data with some regard for the complexity of the data within each class, but with less concern for data accuracy than required for other uses. Recording and communication tasks may best be aided by legibility of the record and by the inclusion of broader concepts and descriptors, with the fine details of the encounter given only when pertinent. The ability to "rapidly retrieve and logically organize and display information within the record"[3] especially when related to a specific clinical problem, is of paramount importance in this setting. Research, audit, and quality improvement efforts, however, may benefit from a finely granular record, controlled data sets and mapping to defined sets of clinical concepts.

MEASURING QUALITY

Using these attributes of a quality medical record -- legibility, accuracy, completeness and meaning -- appropriate outcome variables for a study of the quality of CPRs can be chosen, always keeping in mind that the record should reflect the true state of the patient. Ultimately, it could prove beneficial to be able to take a computer-based record system and judge it in some manner, perhaps with a score, against any other system. And ultimately, the test to be answered is whether or not the record communicates according to its purpose. However, given the enormity of that research challenge, intermediate steps can and must be defined that are directed at evaluation of the separate attributes.

Legibility is a significant issue with computer-based records not only because of unusual abbreviations, but also because the format of a record may significantly alter its ability as a tool of communication. The format of records could be evaluated either by the subjective preference of clinicians or by the ability of clinicians to find information within records.

Correctness and completeness can be judged by the inclusion or exclusion and correctness of data elements. The first problem then becomes one of defining data elements. Defined standards for the necessary elements of a history or physical are bounded by the generality of the HCFA Evaluation and Management categories and the refined sets created for specialized medical fields or problems. It is intuitive and experiential that the most difficult patient encounter to record is that of the general medical patient, who may present with multiple medical problems in varying stages of diagnostic evaluation. The most stringent measure of correctness and completeness in a computer-based record system must therefore be of an encounter with this type of patient.

Possible outcome variables for measuring the meaning of the records can specifically address the ability of the records to communicate based on their purpose. The ability of records to fulfill billing needs, for example, could be judged either by having experienced coders extract the records or by using a computerized coding tool, with comparison of the products from different recording systems.

The ability of the record to fulfill outcomes research needs could be measured by looking at the number of elements of a data set captured by each method of recording. The Health Outcomes Institute,[2] for example, has laid out data sets for a limited number of medical problems. It might be possible to find a patient with a problem for which the Health Outcomes Institute has created a data set, but even if that is the case, clinical encounters rarely are about just one topic, making this approach limited. It has been demonstrated adequately[6-8, 25] that if a defined data set exists, and the record system contains prompted fields for that data, either by templated free text or controlled data entry, then that record is likely to be a more complete record.

The ability of the record to communicate to another clinician the data which that clinician needs could be assessed directly or indirectly. A direct way of assessment is to have clinicians review records of a patient encounter which have been recorded either by free-text, templated free text, or structured data entry, and then to answer questions as to the "true state" of the patient based on those records, or to make problem lists based on those records. An expert panel would be needed to agree on the true problem list, or the true answers to the posed questions.

An indirect way of assessing this function would be to codify the data from each record into medical concepts. An expert panel could determine the concepts for a particular encounter, the chosen encounter could be viewed by the subject physicians, recorded in the CPR, then each of these records could be codified, and scored for accuracy. One might consider use of SNOMED, especially if a computerized coding tool[36] could be obtained. Two problems with this approach would need to be addressed. First, medical concepts do not necessarily convey either the true state of the patient or the information needed by the clinician, discussion of which would merit a separate paper or book. Second, use of a computerized tool would introduce the tool's own inaccuracies as a variable.

For any of these attributes, an intermediate step in the evaluation process is the determination of how much homogeneity or reliability exists within any recording system. That is, given the same input data, will a consistent record be produced by a set of users? Before recording systems can be compared, it must be determined how well any set of records produced using those systems represent the systems as a whole. Completeness and correctness, providing adequate definitions can be determined, are proposed as measures of this consistency.

COMPLETENESS AND CORRECTNESS

Hogan and Wagner[4] have appealed for the use of completeness and correctness in studies of computer-based patient records. Completeness is defined as the proportion of observations in the gold standard that are actually recorded in the CPR; correctness is defined as the proportion of the CPR observations that are correct. Their calculations for these terms are illustrated in Table 1. Two problems arise with this model. First, these calculations of completeness and correctness are not uniformly agreed upon in the medical literature. Second, this model is unfortunately too simplistic for the complexity of errors which can occur in patient records. Elements, unlike test results, are not positive or negative, but present or absent and correct or incorrect.

As an illustration of the first problem, look to the work of Barrie & Marsh[23] and Rickets et al.[27] who used an ideal list of keywords as their gold standard (judging presence vs. absence from this list) against a computerized list of keywords. They measured completeness as the ratio of the number of keywords present to the number that should have been present,

	Gold Standard Positive	Gold Standard Negative	total
CPR Positive	a	b	a+b
CPR Negative	c	d	c+ d
total	a+c	b+d	a+b+c+d

Completeness = a/a+c	Correctness = a/a+b

Table 1: Hogan/Wagner model for completeness and correctness[4]

or (a + b) / (a + c), (see Table 2) that is, including in the numerator both the correct and any incorrect or extra keywords. Correctness was defined as the ratio of the number of keywords present and correct to the total number of keywords, or a / (a + b) consistent with the Hogan/Wagner model. Even Hogan and Wagner did not follow their own model in their earlier work on medication errors, where completeness was really a measure of incompleteness, defined as the mean number of medication records per patient for which no record existed.

Barlow et al.[28] looked for the presence and correctness of five observations in their computerized records, using the paper records as their standard. Their measures of completeness and correctness are not based, however, on the numbers of observations, but rather on the number of completed records,

	Ideal Keyword Present	Ideal Keyword Absent	total
CPR Keyword Present	a	b	a + b
CPR Keyword Absent	c	d	c + d
total	a + c	b + d	a + b + c + d

Completeness = (a + b) / (a + c)	Correctness = a / (a + b)

Table 2: Barrie/Ricketts model for completeness and correctness[23, 27]

some of which may have had more than one observation error. This type of variance in definition makes the comparison of studies difficult.

A comprehensive and mutually exclusive classification of data elements and errors proposed in this study is, as follows:

1. Elements which are present in the gold standard and which are correct (this will be referred to as n1, correct elements);
2. Elements which are present in the gold standard but which are recorded incorrectly (n2, incorrect elements);
3. Elements which are present in the gold standard but which are absent from the record (n3, missing elements); and
4. Elements which are absent from the gold standard but present (and unsubstantiated) in the record (n4, extra elements).

A similar classification for errors was used by Fortinsky & Gutman[19] looking at the reliability of computerized coded problem lists. They measured errors of omission (n3), unsubstantiated problems (n4), and incorrect codes (n2). Results were reported in these values and also as the calculated value of the percent of records which contained no errors.

Using the descriptions for completeness and correctness given by Hogan & Wagner, it appears that completeness must be (n1 + n2) / (n1 + n2 + n3) (the proportion of observations in the gold standard that are actually recorded in the CPR) and correctness must be n1 / (n1 + n2 + n4) (the proportion of the CPR observations that are correct). In attempting to fit this classification (n1 - n4) into Table 1, the question arises as to the appropriate labeling of the gold standard columns. The gold standard in this case is not a test which can be described as positive or negative when repeated over, for instance, 90 different samples. Instead, it is a list with 90 different elements present in that list and an unknown and possibly infinite number of elements absent from it. If, rather than the terms "gold standard positive or negative" the model uses "element present or absent from the gold standard list", then Table 3 could be constructed. Now the calculation of completeness meets the definition above, but the calculation of correctness does not. Correctness would be overrated because of the inclusion of n2 in the numerator.

	GS‡ Element Present	GS‡ Element Absent	total
CPR Present	a = n1+n2	b = n4	a + b = n1+n2+n4
CPR Absent	c = n3	d = NQ*	
total	a + c = n1+n2+n3		
Completeness = a / (a + c) = (n1+n2) / (n1+n2+n3)		Correctness = a / (a + b) = (n1+n2) / (n1+n2+n4)	

* Not Quantifiable ‡Gold Standard

Table 3: Gold standard: present vs. absent

If, however, "gold standard positive or negative" becomes "element matches or does not match the gold standard", then Table 4 could be constructed. This gives an appropriate calculation of correctness, and is the calculation used by Hogan and Wagner in their article on medication errors.[32] It ignores the contribution to completeness of the incorrect elements (n2), however, and therefore does not fit the Hogan/Wagner definition of completeness.

	Element Matches GS‡	Element Does Not Match GS‡	total
CPR Present	a = n1	b = n2+n4	a + b = n1+n2+n4
CPR Absent	c = n3	d = NQ*	
total	a + c = n1+n3		
Completeness = a / (a + c) = n1 / (n1+n3)		Correctness = a / a + b = n1 / (n1+n2+n4)	

*Not Quantifiable ‡Gold Standard

Table 4: Gold standard: match vs. no match

Not all studies of data quality require a measure of incorrect elements; either the element is present or absent in the standard and present or absent in the computer record but without concern for present but incorrect elements. However, for studies such as this one where incorrect elements must be considered, the calculations of completeness and correctness will not conform to those given by Hogan & Wagner. The calculations for completeness and correctness which will be used in this study conform to the definitions used in the Hogan/Wagner model but not to any single table presented here. Completeness is defined as the proportion of elements in the gold standard that are recorded in the CPR, and is calculated as:

$$\text{completeness} = (n1 + n2) \;/\; (n1 + n2 + n3)$$

Correctness is defined as the proportion of elements in the record which are correct as compared to the gold standard, and is calculated as:

$$\text{correctness} = n1 \;/\; (n1 + n2 + n4)$$

THE RESEARCH OBJECTIVE

The purpose of this research is to define a methodology for the evaluation of the quality attributes of completeness and correctness in clinical encounter data as recorded in computer-based patient records. Clinical encounter data is defined as the patient history, physical examination, clinician's impression and plan based on a particular encounter of the clinician with the patient. Considerations for the evaluation of this methodology include the ability to generalize results, the reliability of the data, and the taxonomy of errors which occur in medical records.

METHODOLOGY: OVERALL DESIGN

With completeness and correctness the outcome variables being studied, and in order to judge solely the recording method (the CPR or some portion of it), other confounding variables need to be eliminated, if possible. The patient encounters should be standardized, the subjects should be uniform and have previous knowledge of the system being tested, and the tools for recording within any computer-based system must be chosen appropriately, such that the only remaining variable is the interaction of the clinician with the method itself.

Two videotaped patient encounters were used as the study material. Only the history portions of the encounters were used since verbalization of physical findings by the examining clinician would have been required for the subjects to record a physical examination. The gold standard for data content of the patient encounters was determined by an expert panel consisting of three clinicians with significant clinical experience using an iterative process, with the goal of defining the individual data elements expected in any record which might result from that encounter.

Eight clinicians (seven physicians and one nurse practitioner) were recruited to be the study subjects and were blinded to the purpose of the study. In individual sessions, each clinician was shown both of the videotaped encounters. After viewing each tape, the clinician was asked to record the encounter either by dictation or using computerized forms which allowed for both keyboard entry of free text into templates and structured data entry. Both order of the viewing and method of recording were randomly assigned, but each clinician recorded one encounter by each method. This resulted in eight records for each patient encounter, four by dictation and four using the CPR, for a total of sixteen records.

The records were transcribed or printed, then abstracted by this author looking for the presence and correctness of the elements which were found in the gold standard. A second abstraction by another researcher was performed on a percentage of the records as a check on the reliability of this process. In addition, any data in the record which was not present as an element in the gold standard and was not substantiated by review of the patient encounter material was noted.

METHODOLOGY: PATIENT ENCOUNTERS, the test material

Four videotaped patient encounters were obtained. These were labeled Patients A - D. Patients A and B were actors who were presented as new patients. That they were actors was known to the clinicians who interviewed them on videotape, but significant suspicion of this fact was not raised with the subject clinicians. Patients C and D were actual patients who presented to their own physicians for follow-up of multiple medical problems and who consented to use of the videotaped material for research purposes. Patients B and D were used for the actual study. Patients A and C were used to train the study subjects to the particular tools of the computer-based record system.

Because of the different types of encounters, new patient versus follow-up patient, different computerized forms were best suited for recording the interviews of Patient B and Patient D. Therefore, the training interview with Patient A or C was matched for each subject with the interview for which that subject used the computerized tool, that is, if the subject was to use the CPR for Patient B, then Patient A was used as training; if the Patient D interview was to be recorded on the CPR, then Patient C was used for training. The training tape was viewed before the study tapes regardless of the order of viewing of the subsequent tapes, and the fact that the purpose was for training was not revealed to the subjects. A standard tape recorder was used for capturing the dictations. Because dictation equipment is commonly used, a similar training procedure was not included for the use of the dictation method.

Videotaped patient interviews were chosen as the testing material in this study. Two goals were satisfied with this choice. The first was to provide a testing situation near to real life. The second was to assure that every subject was given identical material to record. In addition, this method allowed the testing material to be portable, which improved the ability to recruit subjects who were prior users of the computer-based record system.

The use of videotaped patient encounters for assessing the quality of medical evaluation and medical records is not unusual.[7, 11, 15, 33, 37, 38] Residency programs often videotape the resident staff encounters in order to critique their skills, and other studies have evaluated record-keeping quality based on videotapes of patient encounters. With the exception of one study using videotapes of upper abdominal ultrasonography,[7] however, the videotapes are used in these studies as a single measurement and not as a tool to be reused with a set of study subjects.

Videotapes were chosen rather than audiotapes in order to more accurately reflect the patient encounter. This proved useful in several instances where the patients motioned to indicate sites of discomfort. With the exception of one blood pressure recording, only the history-obtaining portion of the patient encounters were recorded. Unless the clinicians had chosen to verbalize physical findings, accurate knowledge of the findings could not have been conveyed through the videotapes.

What are the other options?

The gold standard for many studies of data accuracy are paper-based records.[4] We know that paper-based records are inaccurate and incomplete,[11-15, 17, 39] however, and studies using paper-based records as a gold standard are not answering today's correct question. Patient records will be increasingly computerized, whether better or worse than paper records. The appropriate questions to ask must include questions which compel the comparison of computer systems or of data capture methods, or require analysis of data errors or usability issues. In questions of the quality of CPRs, the paper record now becomes unimportant, and the gold standard must be the patient and the patient encounter.

Several other testing methods using patient encounters could potentially have been employed. Standardized patients might have been used. In other studies, they have been successfully integrated into practices with a low rate of detection.[14, 40, 41] While presumably more realistic, this method adds as a confounder the variability in content of the encounter. For that reason, as well as the difficulty of recruiting and training actors, this method was not chosen.

Viewing of the actual patient encounter by one or more observers has also been employed,[12, 17] where the research involves comparison of records made by the observer(s) and by the clinician who participated in the encounter. More than one observer could be used as a check on inter-rater reliability. While this method would be ideal to assure that the testing material is true to life, it is not easily scaled up for use with many-subject studies. If the encounters were staged, then they could be repeated for multiple observers, but this again poses significant technical difficulties.

What are the problems with this method?

Videotapes have primarily been used in the past to study the process, not the content of care. The question arises as to whether or not a videotape accurately enough reflects a patient encounter to use it as a standard. The fallibility of all methods of recording encounters has been well pointed out by Gerbert and associates[11, 37] who looked for the "truth" about the content of patient encounters and concluded that more content was recorded by interview of either the patient or clinician immediately after that encounter than either by the videotape or chart for that visit.

Several considerations arose in preparation and use of this videotaped material. The first problem that was apparent was the need for accurate background information on the patients who were taped in follow-up visits.

Since this information had not been obtained at the time of the encounters, past history consistent with the data on the videotapes was created by this author. It was entered into the computer-based record system as prior visits and therefore was available to the clinicians who recorded on that system just as it would have been in the office setting. For clinicians dictating those encounters, the information on the computer-based system was printed out and given to the subjects prior to viewing the videotapes. One error was made in this process which was not caught until the study period had already started. One patient mentioned taking Zantac, yet this was not on her medication list. It was also not listed as an element in the gold standard list discussed later. Therefore, any mention of it was not scored in the results.

More than one of the study clinicians asked if they could record what they would have done had they been the clinician involved in the encounter, and this must frequently be a temptation when viewing another clinician's work. The content, and not the quality of the patient encounter, was the issue in this study, however, and the study clinicians were reminded of this.

METHODOLOGY: THE SUBJECTS

The ideal for this study was to conduct it with subject clinicians who use the CPR in their regular practices. This proved to be more difficult than expected. Eight clinicians from this community were recruited to be the study subjects.

Five had used the CPR regularly in their practices for from 2 - 15 months. Two were familiar with the CPR but did not use it in practice, and the last used computers regularly but had no prior knowledge of the CPR. For this clinician, a brief review of the tool was given prior to the study period. Seven of the clinicians were in active practice at the time of the study, seven were physicians (two family physicians, three internists, two emergency physicians), one was a family nurse practitioner.

The instruction given to the study subjects was simply to record the patient encounter as viewed, in its entirety, and in a manner similar to the way they might record a real office visit. It was emphasized to them that if data did not fit the form as designed, that it should be recorded nonetheless, that they were not being judged by placement of the information only by its presence. If at any point the subject asked a question concerning the necessity of recording data, the only response given was that if the data was present in the videotape, that it should be recorded. No further information about the purpose of the study was given to the subjects.

The three videotapes, first the training and then the two study tapes, were viewed sequentially, with the encounter data recorded at the end of each tape. Each session took approximately 1 hour per subject and in no case was there significant disruption of the study period.

METHODOLOGY: THE CPR and development of the encounter forms

Logician is a computer-based outpatient record system produced by MedicaLogic, Inc. of Hillsboro, Oregon. It was chosen as the CPR for this study because it provides the ability to capture data in a fully structured fashion, as free text or as templated free text, for the ability to customize forms within the system, and because of the availability of users in this area. For this study a demonstration version of Logician 4.2 was used, which did not have a number of features available on current releases, but which was portable. All data was recorded on a laptop computer, which allowed for control over the forms used and for storage of the data, and did not interfere with the clinicians' production versions.

Logician features include scheduling functions, problem lists, medication lists, allergy lists, and directives, all of which can be updated, in addition to the encounter records. Data can also be aggregated into customized tables. Decision support, alerts, coding and many other functions are built into this system but were not utilized in this study.

Clinical encounter data can be captured in Logician by a number of methods. Three types of encounter forms are available: blank notes, Note Templates, and Encounter Forms. Into any of these, free text can be transcribed after dictation or entered by keyboard. The Note Templates and Encounter Forms by their structure serve as prompts for data. In both, certain data from prior visits can be pulled forward and the use of predefined or "boiler-plate" text is possible. In addition, the templates can be printed for making handwritten notes. The Encounter Forms provide the additional feature of structured data entry, with radio buttons, action buttons, check boxes, drop-down lists, flowsheet views, and single and multiline edit fields. Encounter Forms and Note Templates can be created by the users and a Web site is made available by MedicaLogic where the creators can upload those forms to be shared with other users.

Two Encounter Forms were used for this study (see Appendix A). They do not represent the full capabilities of Logician, but for purposes of this study, further functionality was not necessary. The Multiple SOAP Note is a set of forms which contain multiple structured data fields for physical examination and follow-up instructions but contain no structured data fields appropriate for this study. The fields used were templated to structure but did not contain predefined text. This set of forms was used for Patient D who was seeing her physician for follow-up of multiple medical problems. The General History form is a series adapted by this author from two other form sets and which provides more structured data entry, particularly for the Review of Systems section. It does not contain a physical examination section. Of the 63 elements in the encounter with Patient B, 15 could be recorded in a structured fashion. Space for free text entry was also provided in each section so that users were not constrained to use only the structured data fields.

Ideally, subjects should be allowed to use all available functionality of a given CPR. However, because Logician allows such variety in the manner of data entry, and because the users came with such varied backgrounds in terms of system use, it was felt that any meaningful comparison must depend on restricting use to limited forms. In addition to the Encounter Forms noted above, subjects were allowed to update the Medication, Problem, Prescription and Allergy lists.

The clinician subjects who used Logician in their practices were surveyed as to their practice use of the tool. The number of chart notes created in a week by those clinicians varied from 10 - 150 including telephone notes. None of the subjects reported using the Encounter Forms, and only two ever used the Note Templates, and then only in 10% or less of their chart notes. An average of 84% of charts created by this group were partially or totally dictated notes without use of templates, with the balance recorded by keyboard entry into blank notes.

METHODOLOGY: DATA ELEMENTS, making the gold standard

What constitutes a data element? That is, what is an individual, countable, atomic unit of data in a medical history? An empirical approach was chosen for answering this question based on the belief that an element is significant if it is felt to be significant by clinicians who might make or use the record. This detours a little into the attribute of meaning in the patient record, and away from pure concepts of correctness and completeness, but the author suggests that it is still the best approach. Three physician reviewers including this author, with total clinical experience of 46 years, reviewed the videotapes and in an iterative process determined the "gold standard" for data elements in the videotaped patient encounters (see Figure 1). As a first round, each reviewer was asked to view the videotape of each interview, and to record the encounter, given a form which provided a few prompts and aided in breaking the record into elements. However, each reviewer was specifically instructed not to be restricted by the structure of the form. Each was also allowed to watch the videotape as many times as needed over an unrestricted time frame.

The results from each of the reviewers were then pooled. The lists with the pooled results were returned to the reviewers who were asked to view the tapes another time and to confirm or deny the presence of items in the lists, and also to add any additional items to the lists, if needed. A second

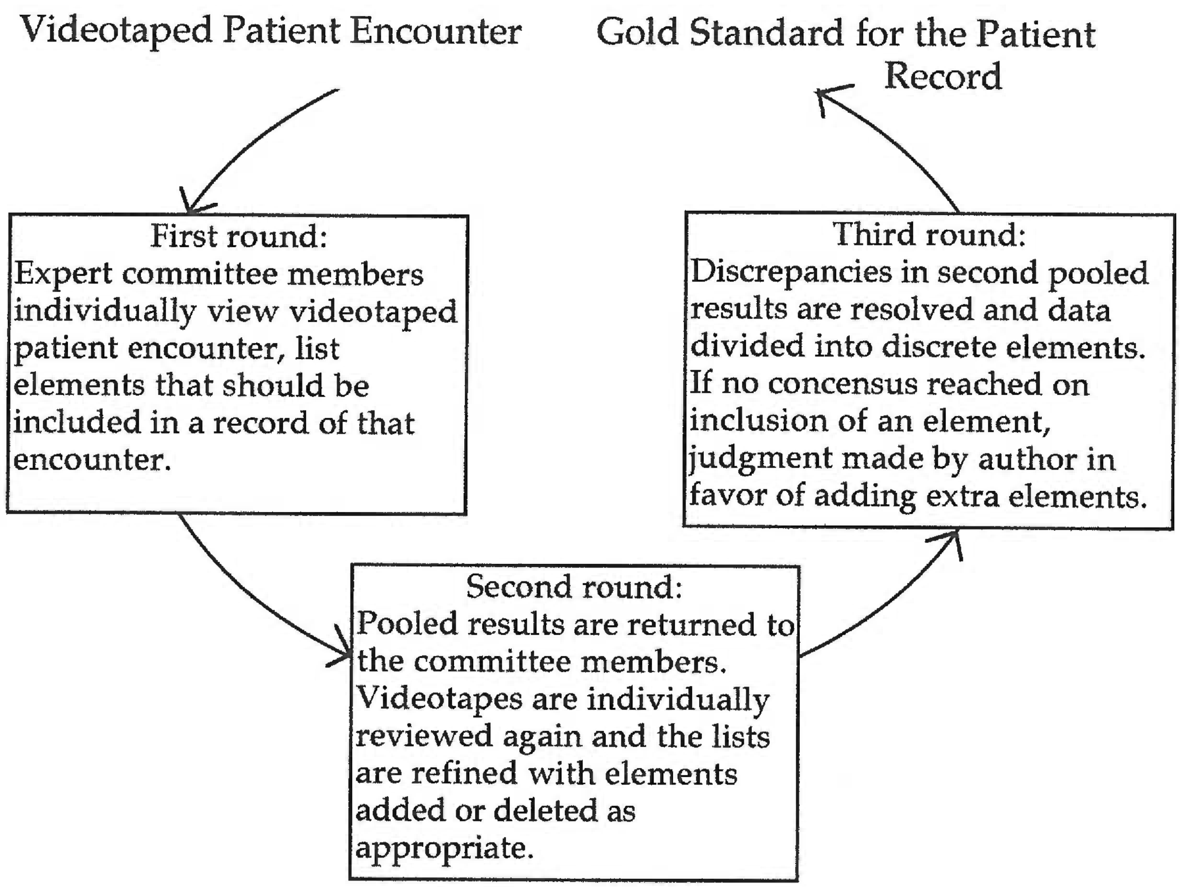


Figure 1: Making the gold standard, an illustration of the iterative process

compilation was performed and checked with the reviewers for discrepancies, at which time they were also asked to divide this data into elements. A question found useful in determining elemental status was: what pieces of data, if taken out of context, add meaning to the record? For example, one patient stated that her mother had died of pancreatic cancer. The family history of pancreatic cancer was felt to be significant, especially in light of her abdominal symptoms. In addition, the fact that her mother had died was felt to be a significant historical fact. Therefore, this was chosen to represent two data elements: mother died, mother had (or family history of) pancreatic cancer. If there was no consensus on an issue this author made the choice, erring on the side of adding extra elements.

What are the other options?

Data elements which may be chosen for study include summary information, such as problem lists,[17, 19, 31] diagnoses,[5, 24, 42] or keywords,[23, 27, 28] specific historical items such as medications[32, 39] or treatments,[26, 30] predefined standards for content of examinations,[6-8, 12, 18, 21, 38, 43-45] or established criteria for evaluation of specific diseases.[11, 20, 37, 45] None of these approaches are appropriate for evaluation of overall documentation over a wide variety of patient encounters. Several approaches have been taken in the past for determination of the units of information to be measured when looking at broader evaluations. Some studies have started with categories of information and abstracted content based on these,[13, 16] (see Tables 5 and 6) while others have started with the content of records and built categories or set criteria based on that content.[14] The categories may be coarse[33] (see Table 7) or refined.[14] As a comparison, the method used in this study resulted in an average of 45 elements per encounter.

Zuckerman[16] used items that are typical of medical audits and a few aspects of medical care not typically audited. (see Table 5) Coding was generous, giving credit to partial or minimal entries, and resulted in 12.7 items per encounter.

Romm and Putnam[13] defined a unit of information as a usual component of a medical encounter (see Table 6). For example, a "respiratory unit" could be the notation of the absence of complaints of cough, sputum production or SOB or the notation of no pulmonary problems. Because of the summary nature of the units of information, the units measured were only an approximation of the degree of agreement. The average number of units per encounter was 13.6.

Norman and associates[14] established the critical information and actions necessary for achievement of acceptable performance for each of their simulated patients. Criteria were suggested by four family physicians and one specialist in the area represented by the simulated case history and then pooled. This resulted in 28 to 54 items per case.

Moran and associates[15] compared the contents, as determined by trained observers, of 22 videotaped patient encounters with the charts resulting from those encounters. While an exact description of the units of comparison is lacking, these units were categorized and subcategorized in a fashion indicating that overall documentation was considered, and gave 14.2 items per encounter. In addition, items were weighted as being very, moderately, or not significant in relation to the complaints stated by the patient.

In a similar manner, Pringle, Ward and Chilvers[33] in their fourth study judged the completeness of the content of computer-based against paper-based records according to the number of "items" each contained (see Table 7). A total of 1195 items were found in the records from 1000 encounters (average 1.2 items per encounter). In a second part to this study, the videotapes from patient encounters were reviewed for number of "topics" they contained and both paper and computer-based records were examined for these topics. Topics were identical to the items of the first phase of the study except for the examination topic. Their rules for determining the presence of a topic were that a topic existed when "both parties use at least one phrase or sentence in its discussion; or if a prescription review takes place without mentioning the underlying topic explicitly." A total of 1097 topics were present in 200 encounters for an average of 5.5 per encounter.

What are the problems with this process?

This process by itself, even before the actual study, brought out many of the problems inherent in the recording of patient data. The task is not as simple

diagnosis
problems
medications-name
mediations - dosage
medications -purpose
medications -effects
medications -side effects
other therapy (diet/wound care)
allergies
followup appointments
diagnostic studies
chief complaint
iatrotropic stimulus
degree of disability
cause of illness
purpose of followup visit

Table 5: Categories of data in the general encounter (Zuckerman[16])

Chief Complaint
Present Illness --
Systems related to the chief complaint:
head
eyes
ears
•
•
•
hematologic
Medical History --
Remaining systems plus:
habits
allergies
disease exposures
current medications
family history
social history
Impression or diagnosis
Tests:
chemistry
hematology
microbiology
radiology
other
Therapy:
medication
diet, exercise
return appt.
consultation, referral

Table 6: Units of the record (Romm and Putnam[13])

symptoms
diagnosis
prescriptions
examination*
site of symptoms
review arranged
numerical finding
duration of symptoms
preventive advice
investigation arranged
previous test results
referral
preventive action
treatment given

*not counted as a "topic"

Table 7: "items" and "topics" in the general encounter (Pringle, et al[33])

as just recording facts, but must deal with the interpretation of or justified inferences drawn from the facts, as well as summative statements about those facts. A total of 90 elements ended up on the gold standard list for Patients B and D. Even with liberal interpretation only 40 (44.4%) of the elements were agreed upon by all three reviewers on the first round, although this was primarily due to the absence of elements by one reviewer. The other two agreed first round on 69 (76.7%) of the elements. If this is the agreement rate between clinicians able to view the tapes in the best of circumstances, how can we expect better with subjects restricted to one viewing of the tapes in a controlled time frame?

METHODOLOGY: DATA ERRORS

In this study a classification scheme for data errors was chosen which would be both exclusive and exhaustive. Records in this study were abstracted by this author for data elements which were then judged against the gold standard to be either 1) present in the gold standard and correct in the record (Correct Element), 2) present in the gold standard but missing from the record (Incorrect/Missing Element), 3) present in both the gold standard and record but substantially different in value in the record (Incorrect/Untrue Element), or 4) not present in the gold standard and unsubstantiated based on the videotape (Extra/Unsubstantiated Element).

Classifying elements as missing from the reports as judged against the gold standard, that is, representing the completeness of the reports, proved easy to determine. Judging accuracy, however, proved to be difficult. What degree of inference is called correct as opposed to untrue? What summative statements overstate or oversimplify the case or combine elements that are appropriately stated separately? The question which was asked for abstraction of data from the reports was: are the elements in the reports equivalent to the elements determined to be "gold standard"? Several rules were laid out to help answer this question.

First, the format in which the data was presented in the reports was not important as long as the researcher clinically could understand its meaning. For example, if information about the family history was reported in the social history section, it was not deemed important as long as the researcher could recognize this as family history.

Second, the term "equivalent" was to be applied liberally. The interpretation of equivalence depends a lot on clinical judgment. A summation of or inference from an element was usually acceptable as long as it was not contrary to other elements. A useful question to ask was found to be: from this record, would I, as a clinician, know that a particular element is present? Precedence for this liberal interpretation is found in another study,[16] where the basis for coding regulations was whether or not the data in the record or patient interview was adequate for a subsequent physician to determine accurately the nature of the care given or to carry out a medical audit.

Third, elements were classified either as:

a. present in the standard and correct in the report (Element Present, n1);

b. present in the standard but not in the report (Missing Element, n2);

c. present in the standard but reported incorrectly (Untrue Element, n3); or

d. present in the report but not substantiated by the gold standard or review of the videotape or transcription (Unsubstantiated Element, n4).

In addition it was possible for data to be present in the report but not to be in the list, but which was substantiated by review of the videotape or transcript. These elements had not been considered significant for inclusion in the gold standard, in which case they were ignored.

The reports were abstracted by this author. A second abstraction was performed by another clinician on 25% of the records, looking for elements meeting the criterias above (Table 8). Neither the videotapes nor transcripts of those tapes were reviewed by this second abstractor, so that the addition of extra data could not be judged. The results were in agreement on 92.2% of the elements. Kappa statistic for inter-rater reliability was found to be .82 overall, .90 for Patient B, and .65 for Patient D.[b] The overall value is consistent with kappa values accepted in other studies.[6, 11, 14, 33, 41]

---

[b] Kappa statistic calculated as follows:

$$\#\ correct\_by\_chance = 125 \times \left(125/180\right) = 86.8$$

$$\#\ incorrect\_by\_chance = 55 \times \left(55/180\right) = 16.8$$

$$chance\_agreement = (86.8 + 16.8)/180 = .576$$

$$observed\_agreement = (48 + 118)/180 = .922$$

$$Kappa = (.922 - .576)/(1 - .576) = .82$$

		Reviewer 2		
		Incorrect	Correct	
Reviewer 1	Incorrect	48	7	55
	Correct	7	118	125
		55	125	180

Table 8: Results of second abstraction of 25% of records for inter-rater reliability: correct (n1) elements vs. incorrect (n2 + n3) elements

RESULTS

Eight subject clinicians recorded each of the two videotaped patient encounters resulting in a total of 16 records, 8 for each encounter, four of those recorded by dictation and four on the computer-based record system. The encounter for Patient B had a total of 63 possible elements as determined by an expert panel, and the encounter for Patient D had a total of 27 possible elements determined in the same manner (see Appendix B). Tables 9 - 12 show the results for each group of clinicians.

How did the recording methods compare?

Table 13 summarizes the results based on the recording method. Figure 2 illustrates the calculations of mean completeness and correctness for the methods. For completeness, the mean ± SE (95% confidence interval) for

Key for tables:
n1 = elements present and correct
n2 = elements present and incorrect
n3 = elements missing
n4 = extra and unsubstantiated elements

Subject	n1	n2	n3	n4
2	56	0	7	0
5	41	1	21	3
6	46	0	17	0
8	55	0	8	0
total	198	1	53	3

Table 9: Results, Group 1
(Patient B + dictation)

Subject	n1	n2	n3	n4
2	16	1	10	1
5	14	0	13	3
6	17	0	10	0
8	25	0	2	0
total	72	1	35	4

Table 10: Results, Group 3
(Patient D + CPR)

Subject	n1	n2	n3	n4
1	36	4	23	2
3	45	0	18	1
4	52	0	11	1
7	43	2	18	4
total	176	6	70	8

Table 11: Results, Group 2
(Patient B + CPR)

Subject	n1	n2	n3	n4
1	18	0	9	1
3	12	0	15	0
4	15	0	12	0
7	16	0	11	0
total	61	0	47	1

Table 12: Results, Group 4
(Patient D + dictation)

dictation was 67.7 ± 12.7 (55.0-80.4%), for the CPR was 69.9 ± 4.5 (59.2 - 80.6%). For correctness, the values for dictation were 98.2 ± 1.2 (95.3 - 101.1%) and for the CPR were 92.6 ± 2.5 (86.6 - 98.5%). Confidence intervals overlap signifying no significant difference between systems in either measure although there was a trend for the CPR to be less correct than dictation.

Method	Overall				Mean Completeness	Mean Correctness
	n1	n2	n3	n4		
dictation	259	1	100	4	66.7%	98.2%
CPR	248	7	105	12	69.9%	92.6%

Table 13: Results by Recording Method

How often were structured data fields used?

For Patient B, 15 elements in each record could have been recorded correctly in structured data fields. Structured data field entries were made 31 times (50% of possible correct uses) in the four records, and in 5 of those instances, the data was recorded incorrectly. In addition, 6 of the 8 unsubstantiated data elements had been recorded in structured fields. For Patient D, no structured data fields were appropriate and yet three of the four unsubstantiated entries were made into structured fields.

What were the errors made?

Table 13 lists the 24 errors made which were either incorrect responses or unsubstantiated elements. The most noticeable error occurred four times, in 50% of the records for that patient. Subject B appeared to be depressed. She had a number of indicators in her encounter pointing to this. Yet nowhere in the encounter was the word *depression* mentioned; in fact, the only statement about her mood was her statement "No, I'm not sad." Yet, four of the eight subjects reported that the patient STATED that she was depressed.

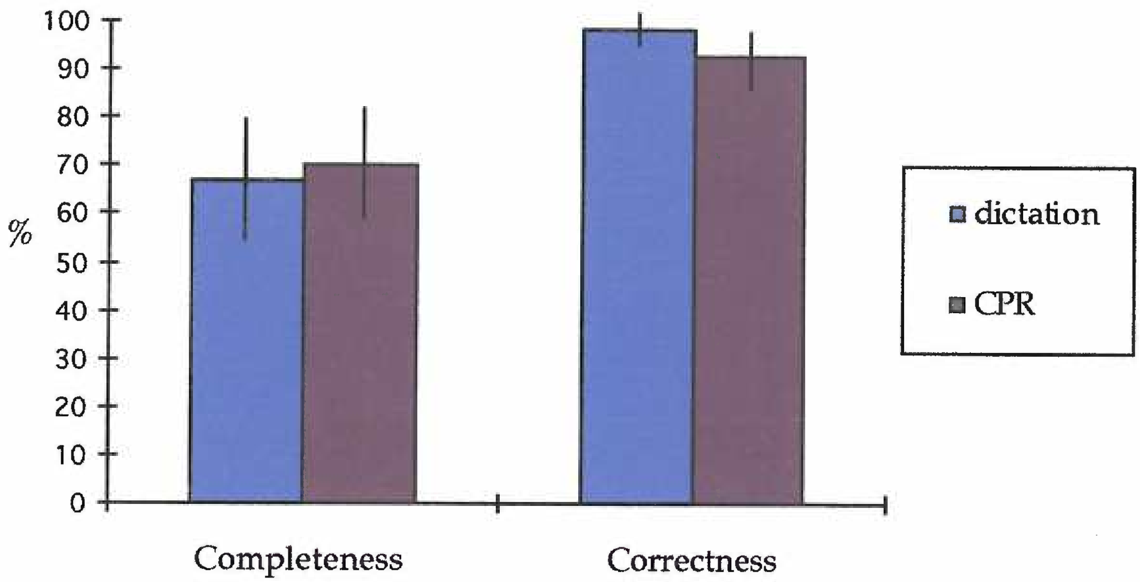


Figure 2: Completeness and correctness in dictated and CPR records, with 95% confidence intervals

Two of these errors occurred using structured data fields, one was by free text entry, and the other was dictated. A fifth subject reported that the patient APPEARED sad, a finding consistent with the encounter.

Three of the incorrect elements were the result of the same clinician checking a series of checkboxes which were titled "Patient denies..." which is positioned to the right of a series of checkboxes which are identical except for the title "Patient complains of...." It is easy to assume that this error occurred because of this juxtaposition and is an example of the importance of layout in structured data charting methods. It is also reasonable to assume that this type of error would occur less frequently with increased familiarity with the recording method.

DISCUSSION

The purpose of this work was to define a methodology for measuring quality in the computerized records of clinical encounters. Definition of this methodology required first the definition of, or more accurately, a dissection of the term quality as it applies to patient records, a choice of the attributes of quality to study, and then a definition of those attributes. Correctness and completeness were chosen as the attributes to be studied for several reasons. First, they are objectively quantifiable. Second, there is precedence in informatics research for the use of these calculations; they have been used in the past as the sole measures of quality, where quality was defined as the product of completeness and correctness. Third, they are a reasonable representation of the internal consistency, the reliability, of a recording method across multiple users.

No statistical significance between the recording methods was either expected or found given this trial study with only eight subjects. The wide variability demonstrated between subjects using the same recording method on the same patient material indicates that a large number of subjects would be necessary to reach statistical significance between recording methods, and even then clinical significance might not be demonstrated. Given the difference in results found between records from the two patients, a variety of

Gold Standard Element	Incorrect Element	Method of data collection		
		Dictated	CPR-free text	CPR - structured
LUQ pain	... and there is no radiation.	X		
GYN examination (presumed to be normal)	... negative based on ...pap smear.	X		
Colonoscopy performed	Add'l Hx: barium enema		X	
S: ear pain	tenderness in neck.		X	
Decreased appetite	Denies: anorexia**			X
possible weight loss	Denies: weight loss**			X
Lack of energy / lethargic / no motivation / tired	Denies: fatigue, malaise**			X
Former smoker (quit 5 years ago)	Tobacco use: no			X
Intermittent (describing pain)	Timing: many times a day			X
Unsubstantiated Elements	**Comments**			
Pain is achy ...	*character of pain was not mentioned*	X		
ROS also is positive for depression.	*see discussion*	X		
...having some claudication	*history of same, but pain not discussed*	X		
Readily admits to feeling depressed	*see discussion*		X	
No fever	*not mentioned*		X	
2.5 mg	*dosage not mentioned*		*	*
Context: when awake	*timing not mentioned*			X
HIV high risk behavior: no (occurred twice)	*not discussed although history of no IVDA documented*			X X
Denies: nausea, vomiting	*not discussed*			X
Pt c/o depression (occurred twice)	*see discussion*			X X
Return to the current provider at a specific date	*followup not discussed*			X
Head: normal	*no exam done*			X
Skin: normal	*no exam done*			X

* unable to determine whether this element was recorded by structured data entry or free text
**these incorrect entries were made by the same clinician and probably reveal a problem with the layout of or unfamiliarity with the structured data entry fields

Table 13: Data Errors

patient encounters might be needed to generalize any results in a definitive study. Fortunately, this method is readily scalable, limited primarily by the ability of clinicians in active practice to contribute their time.

Standards for study methodology

Hogan and Wagner[4] have proposed standards for studies of data accuracy in CPRs, as follows:

1) Report numerical measures of both correctness and completeness;

2) Use an unbiased sampling technique to select patient records for inclusion in the study;

3) Select a gold standard with the intention of approximating the true state of the patient as closely as possible; and

4) Blind the members of the research team who are responsible for the determination of the gold standard to both the purpose of the study and the CPR data when appropriate.

This methodology has either demonstrated or would allow for all of these standards. Numerical measures of completeness and correctness, in accordance with the Hogan/Wagner description but numerically redefined, have been used. Biasing of sampling technique does not apply directly to this methodology but could be approximated by the use of a random sampling of a set of clinical encounters. Until a better definition of the true state of the patient can be determined, then direct viewing of the patient encounter is proposed as an appropriate gold standard. Studies showing that less information is available in a videotape than is actually obtained by the clinician may be faulted by the unwillingness of the researcher viewing the videotape to make assumptions, knowing the purpose of the study. Inherently there is no reason that a videotape cannot represent the entire encounter. Equivalence of the encounter to the true state of the patient is, however, a reasonable question which cannot be answered in this work. The research team responsible for the gold standard was not blinded to the purpose of this study although they were blinded to the CPR data, but this can easily be changed should the methodology be used in a definitive study.

Taxonomy of errors

Even without the ability to attain clinical or statistical significance, value can be obtained from this methodology through application of the errors noted to the development of a taxonomy of errors. There is little in the medical literature about such a taxonomy. Studies on the variability and reliability of medical data[46-48] have not addressed the particular issue of recording errors, nor are lessons from studies on errors in databases appropriate for this topic.

A model of the process of record-keeping suggested by Spackman[49] may provide a useful framework for classification of the errors found in this study. This model divides the process of creation of the content of a patient record into three stages. First, from the encounter certain items are believed to be true by the clinician. The clinician then chooses to record a portion of the items believed to be true. For each item recorded, choices are then made as to the manner of the recording.

Can examples be found that illustrate errors in what the clinicians believed to be true? The most prominent example is that of the statement of depression previously discussed. The nonverbal expressions from this patient and inferences drawn from statements made were apparently so strong that half of the clinicians not only believed that this patient was depressed (a correct statement) but also believed that the patient stated that she was depressed (an incorrect statement). The inclusion of pap smear results based on the history of having had a gynecological examination is an unjustified inference; although more than likely she did have a pap smear at the time of her examination, incorrect documentation of that test could be a clinically significant error. Other examples included the substitution of UGI for upper endoscopy and of barium enema for colonoscopy, and the timing of pain while awake when there was no mention of diurnal variation. Perhaps this set of errors is divisible into the subsets, a) errors of substitution, b) errors of inference, and c) errors of inclusion.

Can examples be found that illustrate errors in what the clinicians chose to record? A measure of this type of error would be similar to the measure of missing elements. What cannot be counted, however, are the elements which the clinicians did not believe to be true and therefore did not choose to record based on that lack of belief. At a maximum, the errors of this type are the same as the missing elements. While it might be assumed that these elements were not recorded because they were not clinically significant, the defined methodology should weaken this assumption. If the gold standard elements are defined by clinicians who are blinded to the study purpose, this assumption should be precluded.

Can examples be found which illustrate errors based on the method of recording of those elements believed to be true and chosen for recording? This is where lessons can be learned in the design of structured data entry forms. Three incorrect entries made by the same clinician are probably errors of this type: instead of checking boxes in the "Patient complains of..." column, the boxes were checked in the adjacent "Patient denies...." column.

A number of errors cannot easily be classified. Is the inclusion of physical findings when no examination was performed an error in inference (external inspection could be considered normal but not other aspects of the physical examination) or in the method of recording (did the presence of checkboxes for normal findings encourage their erroneous use)? Is the description of an ex-smoker by "tobacco use: no" an error in what the clinician believed to be true or an error caused by restriction of choices in structured data fields without a clear definition of those choices? The declaration of no fever when presence fever was not mentioned may be an example of the assumption that if a symptom is not mentioned, then it is negative. The actual value of this element is unknown.

Limitations of this methodology

Two limitations are obvious in this study, one a practical one in the study itself, the other a theoretical limitation with the methodology. An attempt was made in this study to recruit study subjects who were already familiar with the templated and structured data capture methods of Logician. However, not all subjects used Logician regularly, and none of the subjects used its structured data capabilities. How much the variability of the results is a reflection of this lack of familiarity with the recording method is uncertain. The training videotapes were used to try to decrease this confounding variable, but for any definitive study using this methodology, the variability of user training must be considered. In addition, given that the clinician subjects were not necessarily representative of the defined population of clinicians, then any results found here are only estimates of what might be found in a more definitive study.

No attempt is made to claim that measures of completeness and correctness can quantify the concept of quality. Just as the concept of state of the patient is the ephemeral entity represented by the clinical encounter, so quality is only partially and momentarily measured by completeness and correctness. We do not know the value of detail vs. overview in medical records, the value of unasked and unanswered questions, the cost of missing data, or whether or not accurate diagnoses can be reached and decisions made despite disagreement over historical and examination data. What is claimed in this study is that completeness and correctness should be measured in a consistent manner and remain an essential part of the evaluation of quality in computer-based patient records.

CONCLUSIONS

The methodology described in this report can meet the standards for studies of data accuracy in CPRs as recommended by Hogan and Wagner.[4] While not all of their criteria were met in this trial study, adaptation of the method to those standards could be performed without significant difficulty. The greatest limitation to use of this methodology lies in the difficulty of recruiting enough subjects to demonstrate statistically significant differences in recording methods, given the variability between clinicians in recording practices.

The measures of completeness and correctness may not completely define quality in patient records, but are an essential part of that definition. Suggestions have been made for standardizing the definitions of completeness and correctness for studies where both presence and accuracy of data elements is important. In further studies using this method, even if comparative conclusions cannot be drawn, much can be learned about the process of creating a patient clinical encounter record and of the sources of error in those records as produced by the method of recording.

REFERENCES:

1. Feinstein, A. *Clinical Judgment*, (Williams & Wilkins Co., Baltimore, 1967).

2. Dolin, R.H. Outcome Analysis: Considerations for an Electronic Health Record. *MD Computing* **14**, 50-56 (1997).

3. Elson, R.B. & Connelly, D.P. Computerized Patient Records in Primary Care: Their role in Mediating Guideline-Driven Physician Behavior Change. *Archives of Family Medicine* **4**, 698-705 (1995).

4. Hogan, W.R. & Wagner, M.M. Accuracy of Data in Computer-based Patient Records. *JAMIA* **4**, 342-355 (1997).

5. Block, B. & Brennan, J.A. Reliability of Morbidity Data in a Computerized Medical Record System. in *Proceedings of the AAMSI Congress 89* (ed. Hammond, W.E.) 21-30 , 1989).

6. Duggan, A.K., Starfield, B. & DeAngelis, C. Structured Encounter Form: The Impact on Provider Performance and Recording of Well-Child Care. *Pediatrics* **1990**(1990).

7. Kuhn, K. *et al.* Structured Reporting of Medical Findings: Evaluation of a System in Gastroenterology. *Methods of Informatics in Medicine* **31**, 268-274 (1992).

8. Gouuveia-Oliveira, A. *et al.* Longitudinal Comparative Study on the Influence of Computers on Reporting of Clinical Data. *Endoscopy* **23**, 334-337 (1991).

9. ASTM Committee E-31 on Computerized Systems. Standard Guide for Content and Structure of the Computer-Based Patient Record. (1996).

10. Sanazaro, P.J. Measurement of Physicians' Performance Using Existing Techniques. *Western Journal of Medicine* **133**, 81-88 (1980).

11. Gerbert, B. & Hargreaves, W.A. Measuring Physician Behavior. *Medical Care* **24**, 838-847 (1986).

12. Long, T. & Rigers, K. Experience and Reason - Briefly Recorded. *Pediatrics* **56**, 295-296 (1975).

13. Romm, F.J. & Putnam, S.M. The Validity of the Medical Record. *Medical Care* **19**, 310-315 (1981).

14. Norman, G.R., Neufeld, V.R., Walsh, A., Woodward, C.A. & McConvey, G.A. Measuring Physicians' Performances by Using Simulated Patients. *Journal of Medical Education* **60**, 925-934 (1985).

15. Moran, M.T., Wiser, T.H., Nanda, J. & Gross, H. Measuring Medical Residents' Chart-Documentation Practices. *Journal of Medical Education* **63**, 859-865 (1988).

16. Zuckerman, A.E., Starfield, B., Hochreiter, C. & Kovasznay, B. Validating the Content of Pediatric Outpatient Medical Records by Means of Tape-Recording Doctor-Patient Encounters. *Pediatrics* **56**, 407-411 (1975).

17. Bentsen, B.G. The Accuracy of Recording Patient Problems in Family Practice. *Journal of Medical Education* **51**, 311-316 (1976).

18. Jelovsek, F. & Hammond, W. Formal Error Rate in a Computerized Obstetric Medical Record. *Methods of Informatics in Medicine* **17**, 151-157 (1978).

19. Fortinsky, R.H. & Gutman, J.D. A Two-Phased Study of the Reliability of Computerized Morbidity Data. *The Journal of Family Practice* **13**, 229-235 (1981).

20. Jones, R.B. & Hedley, A.J. A Computer in the Diabetic Clinic. *Practical Diabetes* **3**, 295-296 (1986).

21. Maresh, M., Dawson, A.M. & Beard, R.W. Assessment of an on-line computerized perinatal data collection and information system. *British Journal of Obstetrics and Gynecology* **93**, 1239-1245 (1986).

22. Dambro, M.R. & Weiss, B.D. Assessing the Quality of Data Entry in a Computerized Medical Records System. *Journal of Medical Systems* **12**, 181-187 (1988).

23. Barrie, J.L. & Marsh, D.R. Quality of Data in the Manchester Orthopaedic Database. *BMJ* **304**, 159-162 (1992).

24. Johnson, N., Manta, D., Jones, L. & Randall, T. Use of Computerized General Practice Data for Population Surveillance: Comparative Study of Influenza Data. *BMJ* **302**, 763-765 (1991).

25. Kuhn, K. *et al.* The Quality of Gastroenterological Reports Based on Free Text Dictation: An Evaluation in Endoscopy and Ultrasonography. *Endoscopy* **23**, 262-264 (1991).

26. Payne, T., Kanvik, S., Seward, R. & Beeman, D. Development and Validation of an Immunization Tracking System in a Large Health Maintenance Organization. *American Journal of Preventative Medicine* **9**, 96-100 (1993).

27. Ricketts, D., Newey, M., Hitchin, D. & Fowler, S. Markers of Data Quality in Computer Audit: the Manchester Orthopaedic Database. *Annals of the Royal College of Surgeons of England* **75**, 393-396 (1993).

28. Barlow, I.W., Flynn, N.A.K. & Britton, J.M. The Basingstoke Orthopaedic Database: A High Quality Accurate Information System for Audit. *Annals of the Royal College of Surgeons of England* **76 (Supplement)**, 285-287 (1994).

29. Hohnloser, J.H., Fischer, M.R.G., Konig, A. & Emmerich, B. Data Quality in Computerized Patient Records. *International Journal of Clinical Monitoring and Computing* **11**, 233-240 (1994).

30. Wilton, R. & Pennisi, A.J. Evaluating the Accuracy of Transcribed Computer-Stored Immunization Data. *Pediatrics* **94**, 902-906 (1994).

31. Yarnall, K.S.H., Michener, J.L., Broadhead, W.E., Hammond, W.E. & Tse, C.-K.J. Computer-Prompted Diagnostic Codes. *The Journal of Family Practice* **40**, 257-262 (1995).

32. Wagner, M.M. & Hogan, W.R. The Accuracy of Medication Data in an Outpatient Electronic Medical Record. *JAMIA* **3**, 234-243 (1996).

33. Pringle, M., Ward, P. & Chilvers, C. Assessment of the Completeness and Accuracy of Computer Medical Records in Four Practices Committed to Recording Data on Computer. *British Journal of General Practice* **45**, 537-541 (1995).

34. Institute of Medicine. *The Computer-Based Patient Record. An Essential Technology for Health Care*, (National Academy Press, Washington, DC, 1997).

35. van Bemmel, J.H. & Musen, M.A. *Handbook of Medical Informatics*, (Springer-Verlag, Heidelburg, Germany, 1997).

36. Campbell, J.R. *et al.* Phase II Evaluation of Clinical Coding Schemes: Completeness, Taxonomy, Mapping, Definitions, and Clarity. *JAMIA* **4**, 238-251 (1997).

37. Gerbert, B., Stone, G., Stulbarg, M., Gullion, D.S. & Greenfield, S. Agreement Among Physician Assessment Methods. Searching for the Truth Among Fallible Methods. *Medial Care* **26**, 519-535 (1988).

38. McCormick, D.P. *et al.* Use of Videotaping to Evaluate Pediatric Resident Performance of Health Supervision Examinations of Infants. *Pediatrics* **92**, 116-120 (1993).

39. Hulka, B.S., Kupper, L.L., Cassel, J.C., Efird, R.L. & Burdette, J.A. Medication Use and Misuse: Physician-Patient Discrepancies. *Journal of Chronic Diseases* **28**, 7-21 (1975).

40. Russsell, N.K., Boekeloo, B.O., Rafi, I.Z. & Rabin, D.L. Using Unannounced Simulated Patients to Evaluate Sexual Risk Assessment and Risk Reduction Skills of Practicing Physicians. *Academic Medicine* **66**, S37-S42 (1991).

41. Woodward, C.A., McConvey, G.A., Neufeld, V., Norman, G.R. & Walsh, A. Measurement of Physician Performance by Standardized Patients. *Medical Care* **23**, 1019-1027 (1985).

42. Studney, D.R. & Hakstian, A.R. A Comparison of Medical Record with Billing Diagnostic Information Associated with Ambulatory Medical Care. *American Journal of Public Health* **71**, 145-149 (1981).

43. Mayefsky, J.H. & Foye, H.R. Use of a Chart Audit: Teaching Well Child Care to Paediatric House Officers. *Medical Education* **27**, 170-174 (1993).

44. Linn, B.S., Linn, M.W., Greenwald, S.R. & Gurel, L. Validity of Impairment Ratings Made from Medical Records and from Personal Knowledge. *Medical Care* **12**, 363-368 (1974).

45. Thompson, H.C. & Osborne, C.E. Office Records in the Evaluation of Quality of Care. *Medical Care* **14**, 294-304 (1976).

46. Komaroff, A. The Variability and Inaccuracy of Medical Data. *Proceedings of the IEEE* **67**, 1196-1207 (1979).

47. Koran, L. The Reliability of Clinical Methods , Data and Judgments (First of Two Parts). *The New England Journal of Medicine* **293**, 642-646 (1975).

48. Koran, L. The Reliability of Clinical Methods, Data and Judgments (Second of Two Parts). *The New England Journal of Medicine* **14**(1975).

49. Spackman, K.A. Personal communication.

Appendix A: Encounter Forms

The Logician Encounter Forms used for this study.

General History

History of Present Illness 63

Review of Systems -1 64 - 65

Review of Systems - 2 66 - 67

Review of Systems - 3 68 - 69

Review of Systems - 4 70

Review of Systems - 5 71 - 72

Family History 73

Past Medical History 74

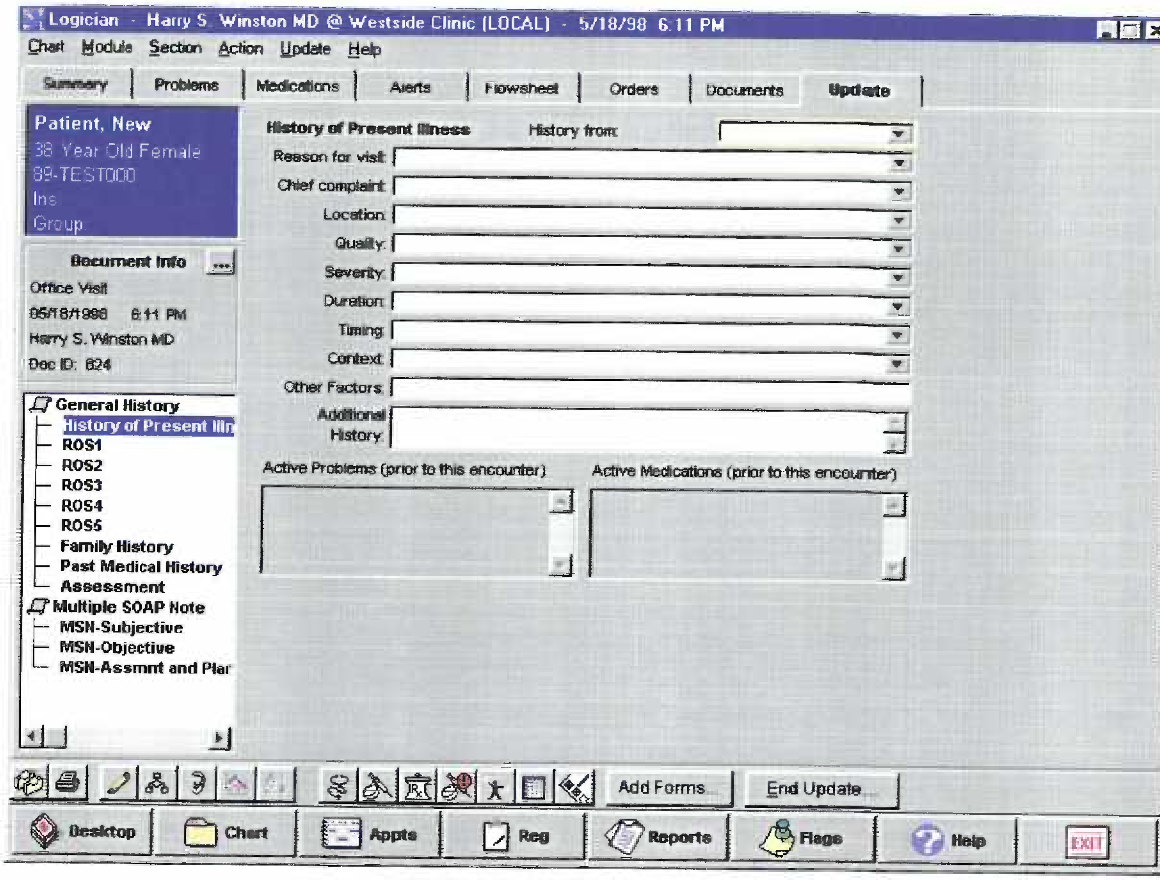
Assessment 75

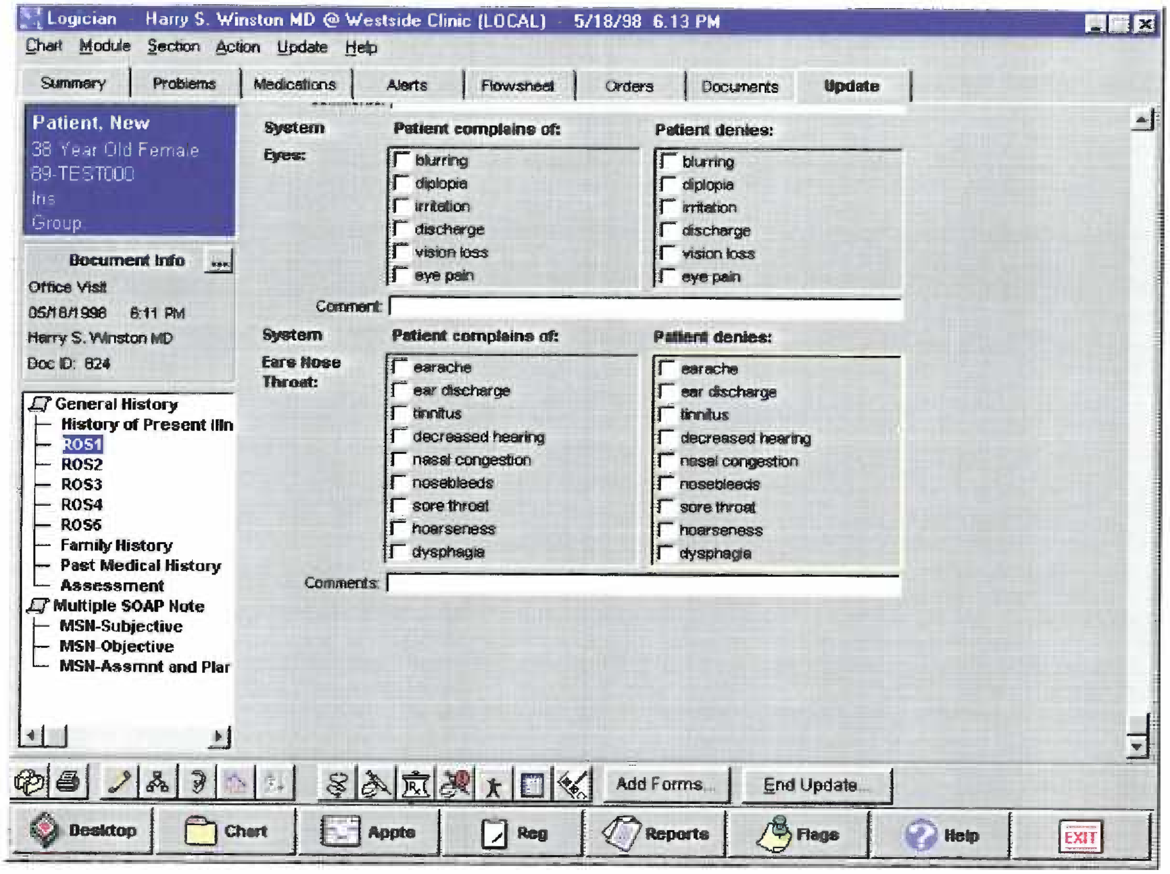
Multiple SOAP Note

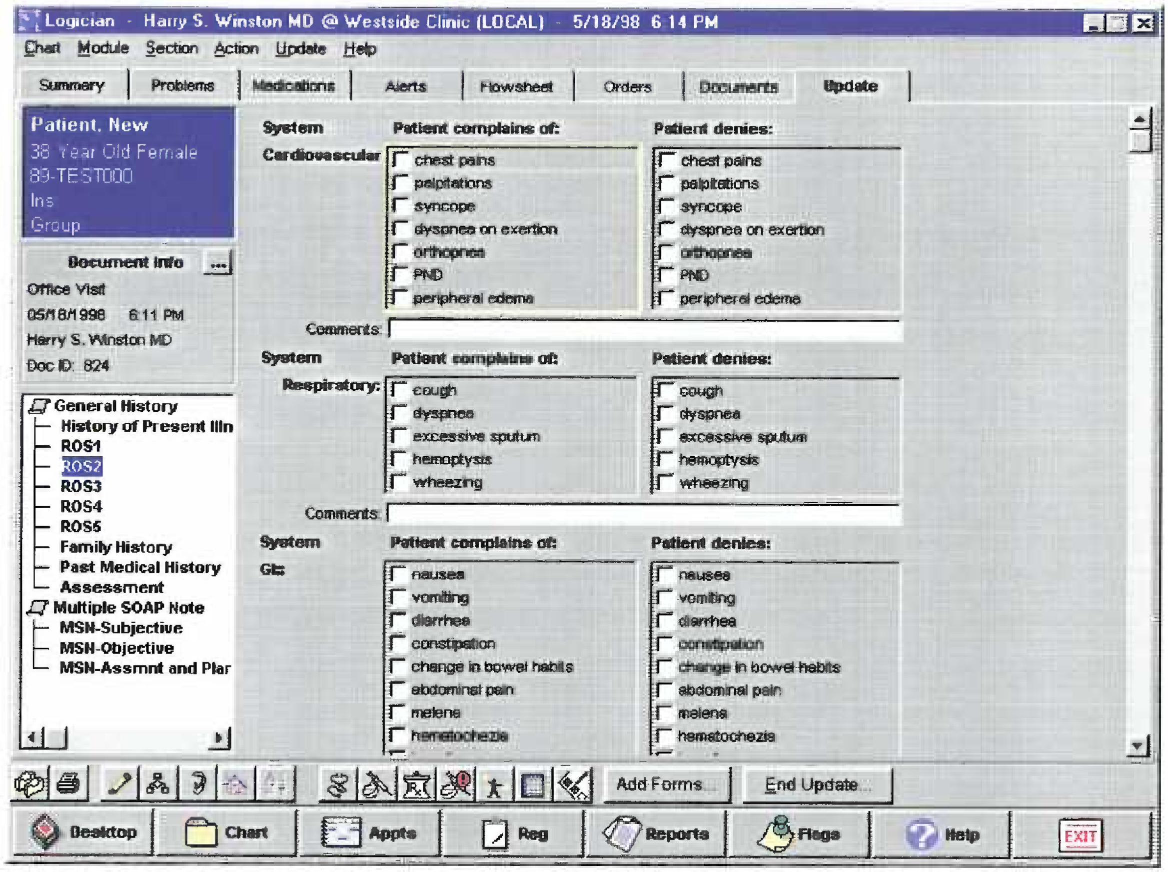
Subjective 76

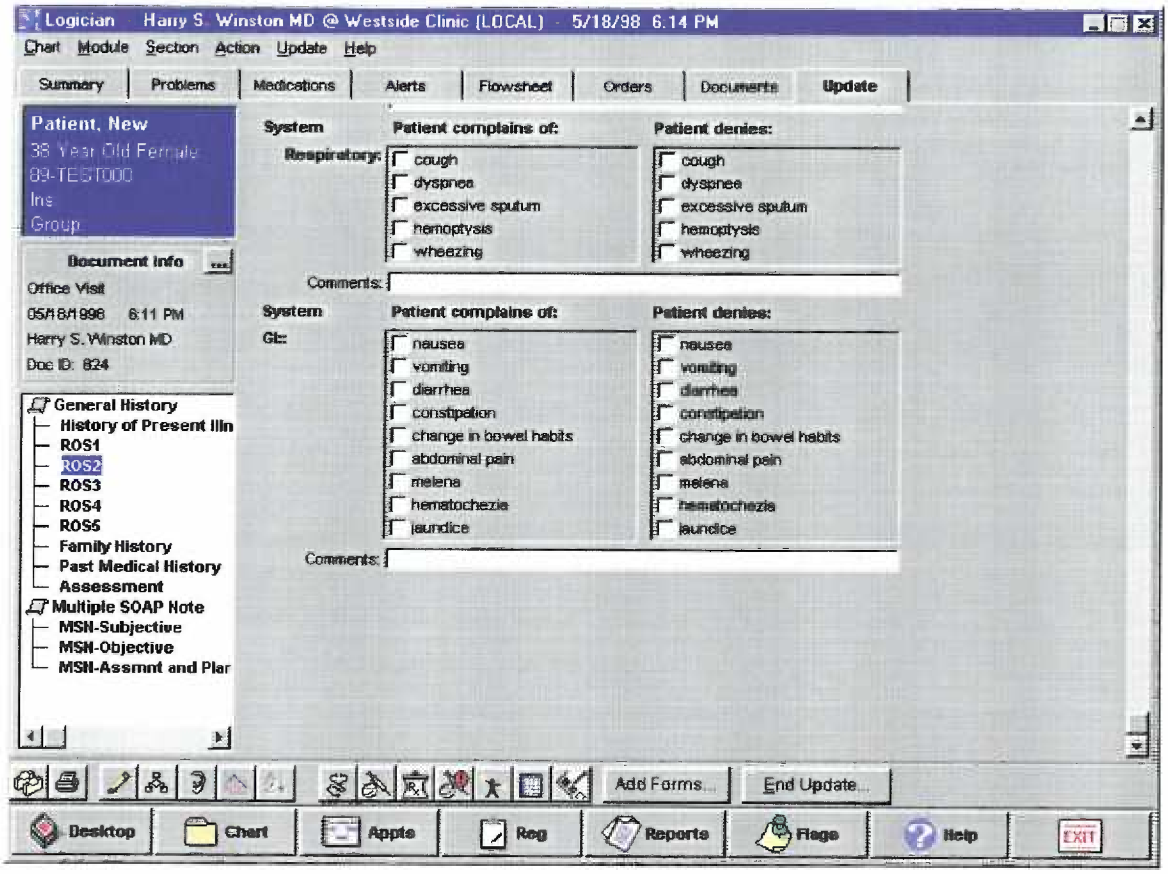
Objective 77

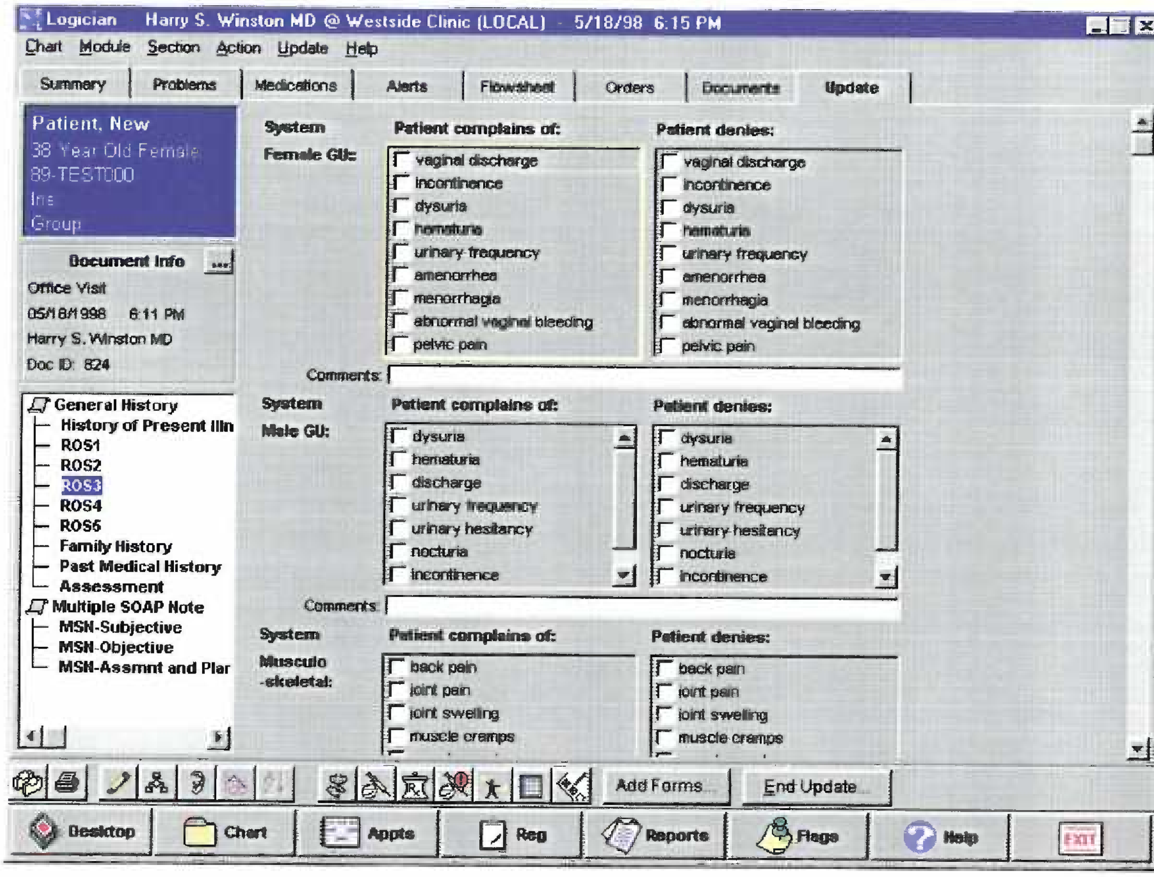
Assessment and Plan 78


Logician - Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6:11 PM
Chart Module Section Action Update Help
Summary
Problems
Medications
Alerts
Flowsheet
Orders
Documents
Update
Patient, New
38 Year Old Female
89-TEST000
Ins
Group
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plan
History of Present Illness
History from:
Reason for visit:
Chief complaint:
Location:
Quality:
Severity:
Duration:
Timing:
Context:
Other Factors:
Additional History:
Active Problems (prior to this encounter)
Active Medications (prior to this encounter)
Add Forms...
End Update...
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT

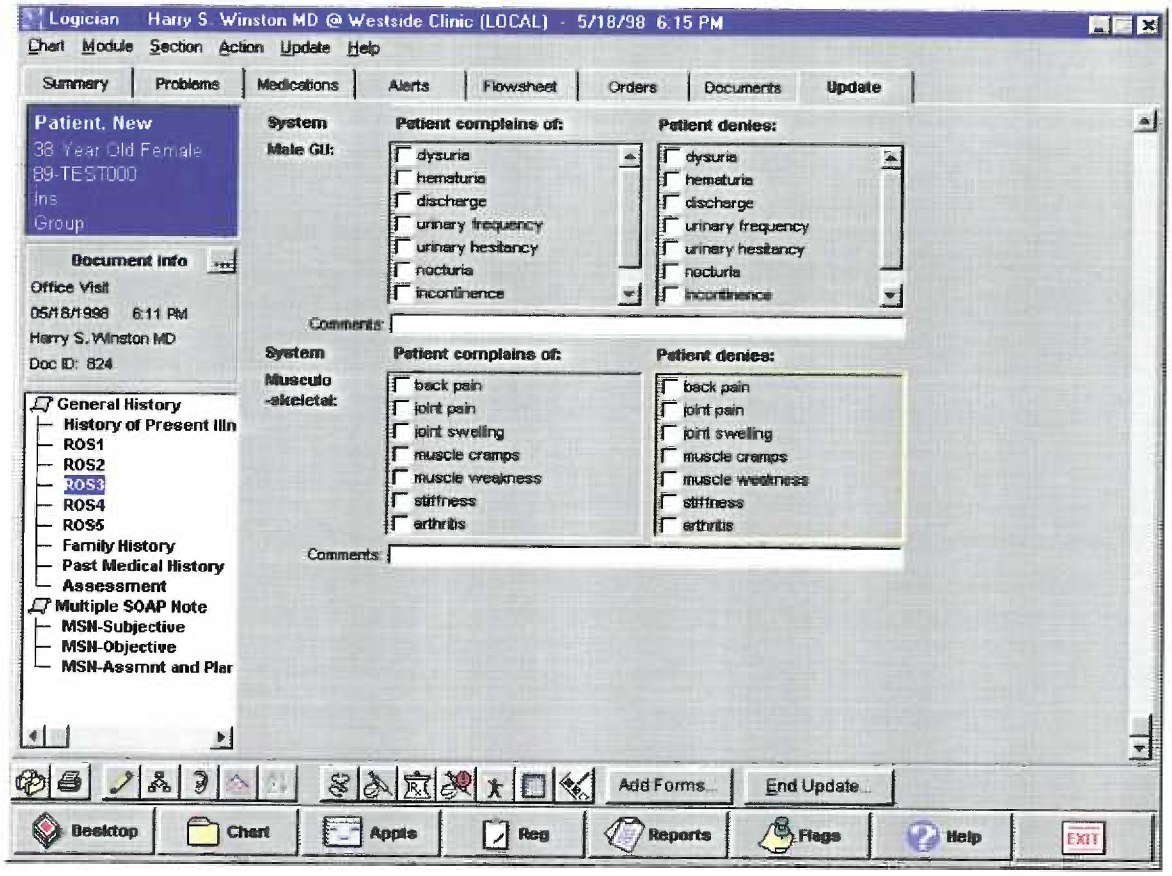

Logician - Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6.13 PM
Chart Module Section Action Update Help
Summary
Problems
Medications
Alerts
Flowsheet
Orders
Documents
Update
Patient, New
38 Year Old Female
89-TEST000
Ins
Group
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 624
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
System
Patient complains of:
Patient denies:
Eyes:
blurring
diplopia
irritation
discharge
vision loss
eye pain
Comment:
Ears Nose Throat:
earache
ear discharge
tinnitus
decreased hearing
nasal congestion
nosebleeds
sore throat
hoarseness
dysphagia
Comments:
Add Forms...
End Update...
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT

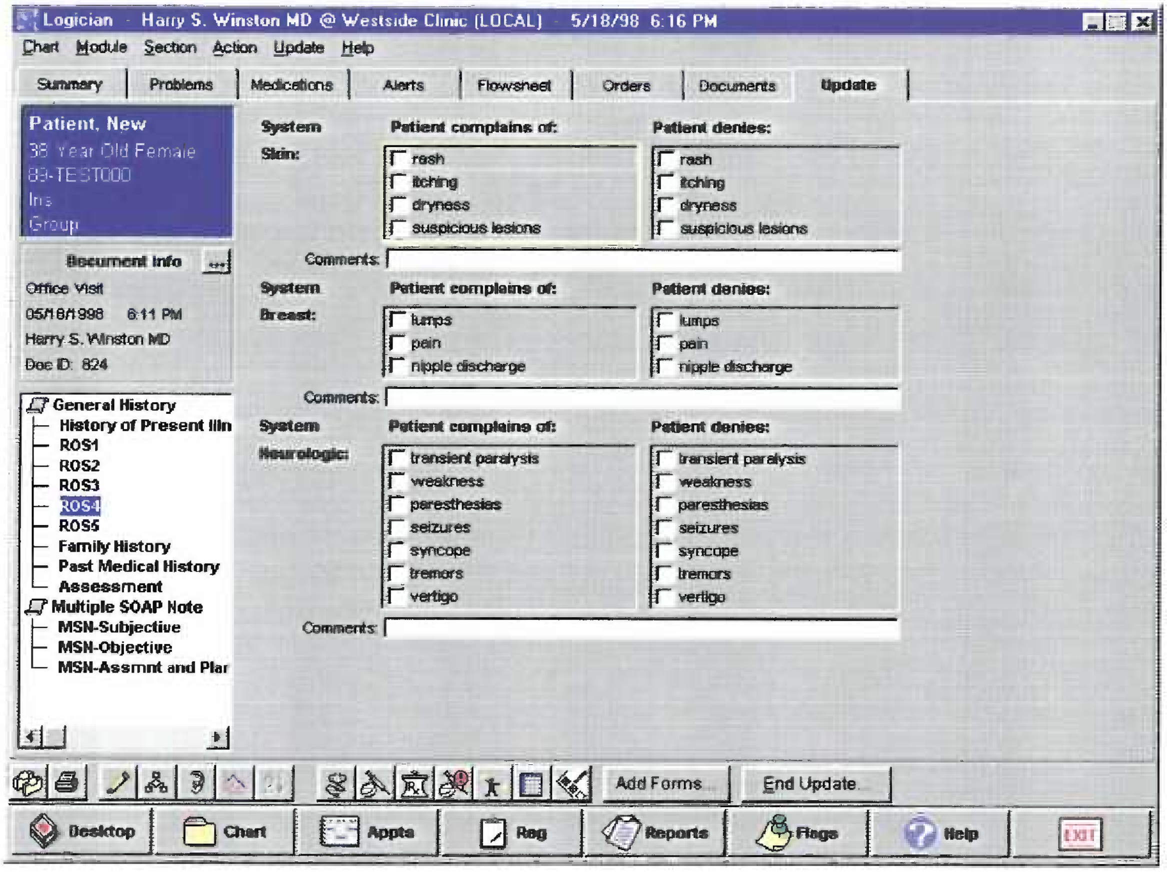

Logician - Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6:14 PM
Chart Module Section Action Update Help
Summary
Problems
Medications
Alerts
Flowsheet
Orders
Documents
Update
Patient, New
38 Year Old Female
89-TEST000
Ins
Group
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
System
Cardiovascular
Patient complains of:
chest pains
palpitations
syncope
dyspnea on exertion
orthopnea
PND
peripheral edema
Patient denies:
chest pains
palpitations
syncope
dyspnea on exertion
orthopnea
PND
peripheral edema
Comments:
System
Respiratory:
Patient complains of:
cough
dyspnea
excessive sputum
hemoptysis
wheezing
Patient denies:
cough
dyspnea
excessive sputum
hemoptysis
wheezing
Comments:
System
GI:
Patient complains of:
nausea
vomiting
diarrhea
constipation
change in bowel habits
abdominal pain
melena
hematochezia
Patient denies:
nausea
vomiting
diarrhea
constipation
change in bowel habits
abdominal pain
melena
hematochezia
Add Forms
End Update
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT

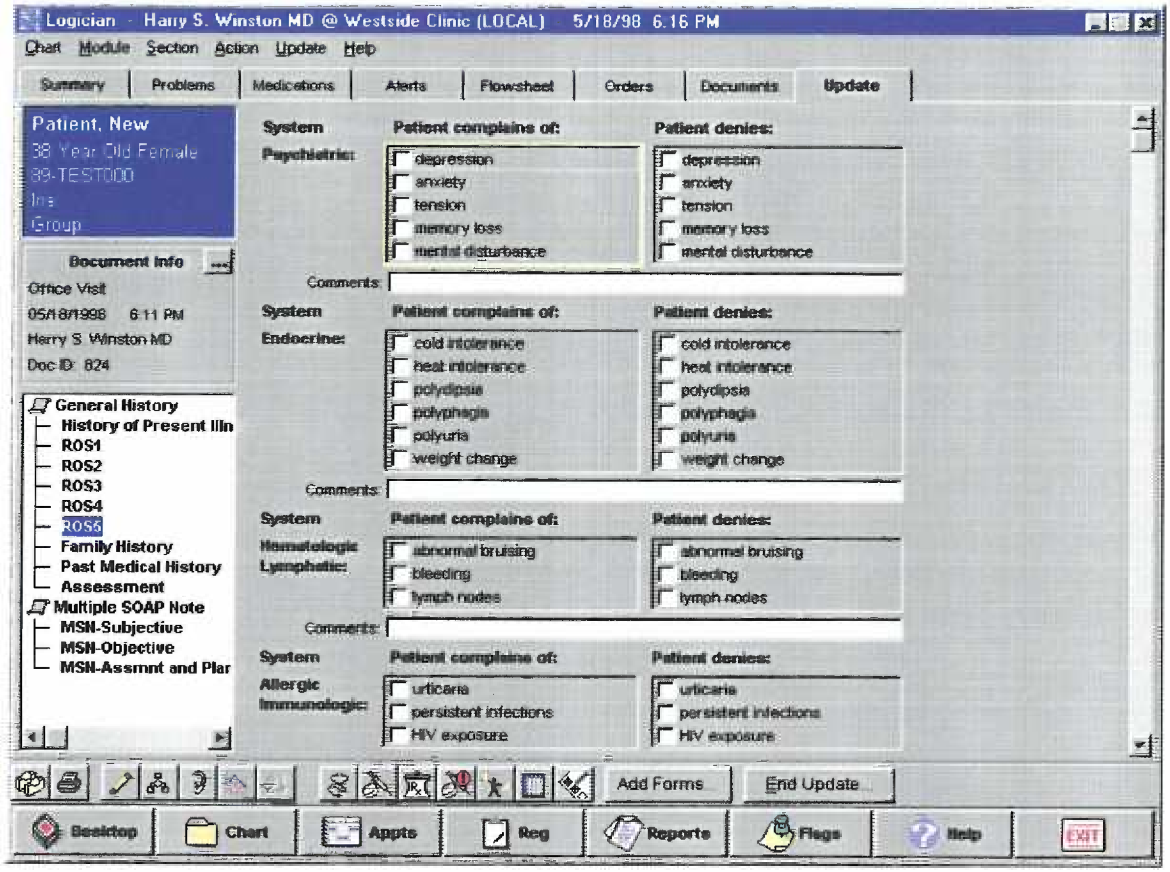

Logician Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6:14 PM
Chart Module Section Action Update Help
Summary Problems Medications Alerts Flowsheet Orders Documents Update
Patient, New
38 Year Old Female
89-TEST000
Ins
Group
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
System
Patient complains of:
Patient denies:
Respiratory:
cough
dyspnea
excessive sputum
hemoptysis
wheezing
Comments:
GI:
nausea
vomiting
diarrhea
constipation
change in bowel habits
abdominal pain
melena
hematochezia
jaundice
Add Forms...
End Update...
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT



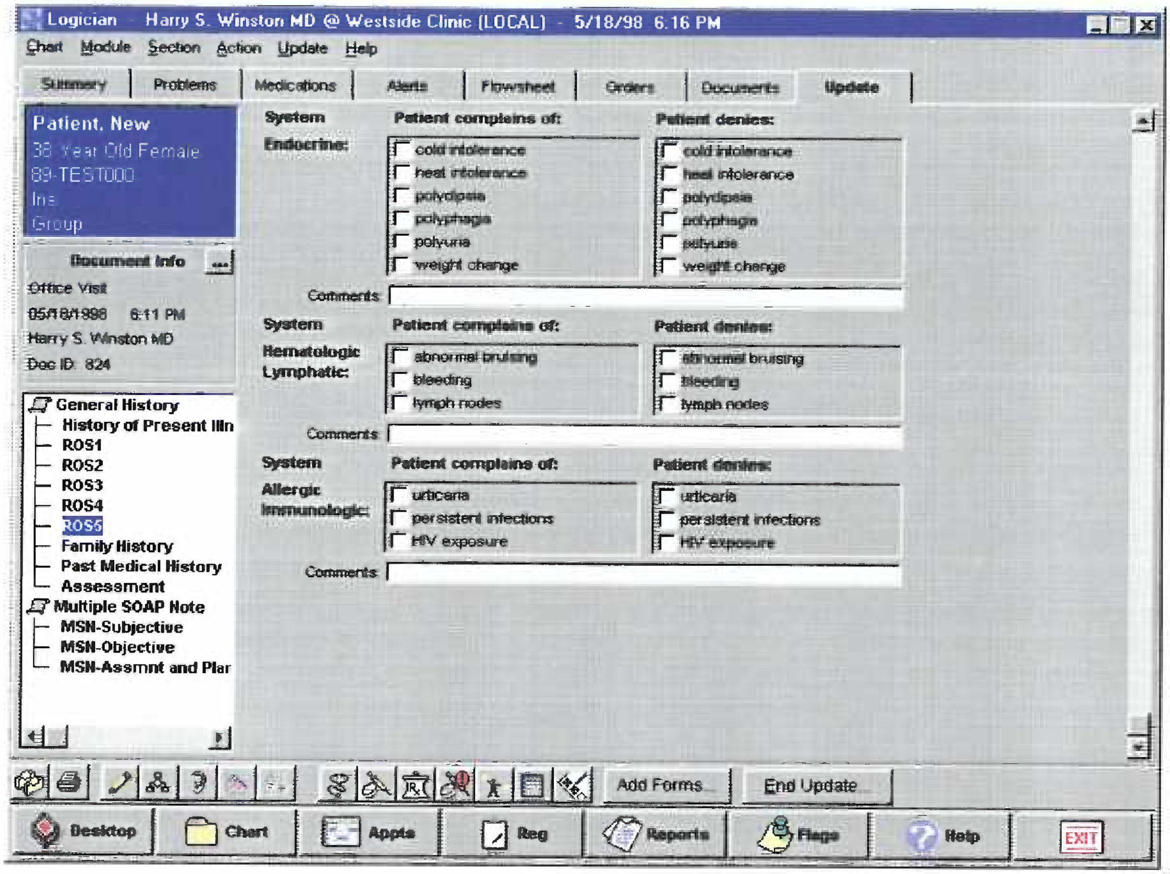
Logician Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6:15 PM
Chart Module Section Action Update Help
Summary
Problems
Medications
Alerts
Flowsheet
Orders
Documents
Update
Patient, New
38 Year Old Female
89-TEST000
Ins
Group
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS6
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
System
Female GU:
Patient complains of:
vaginal discharge
incontinence
dysuria
hematuria
urinary frequency
amenorrhea
menorrhagia
abnormal vaginal bleeding
pelvic pain
Patient denies:
vaginal discharge
incontinence
dysuria
hematuria
urinary frequency
amenorrhea
menorrhagia
abnormal vaginal bleeding
pelvic pain
Comments:
System
Male GU:
Patient complains of:
dysuria
hematuria
discharge
urinary frequency
urinary hesitancy
nocturia
incontinence
Patient denies:
dysuria
hematuria
discharge
urinary frequency
urinary hesitancy
nocturia
incontinence
Comments:
System
Musculo -skeletal:
Patient complains of:
back pain
joint pain
joint swelling
muscle cramps
Patient denies:
back pain
joint pain
joint swelling
muscle cramps
Add Forms...
End Update...
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT

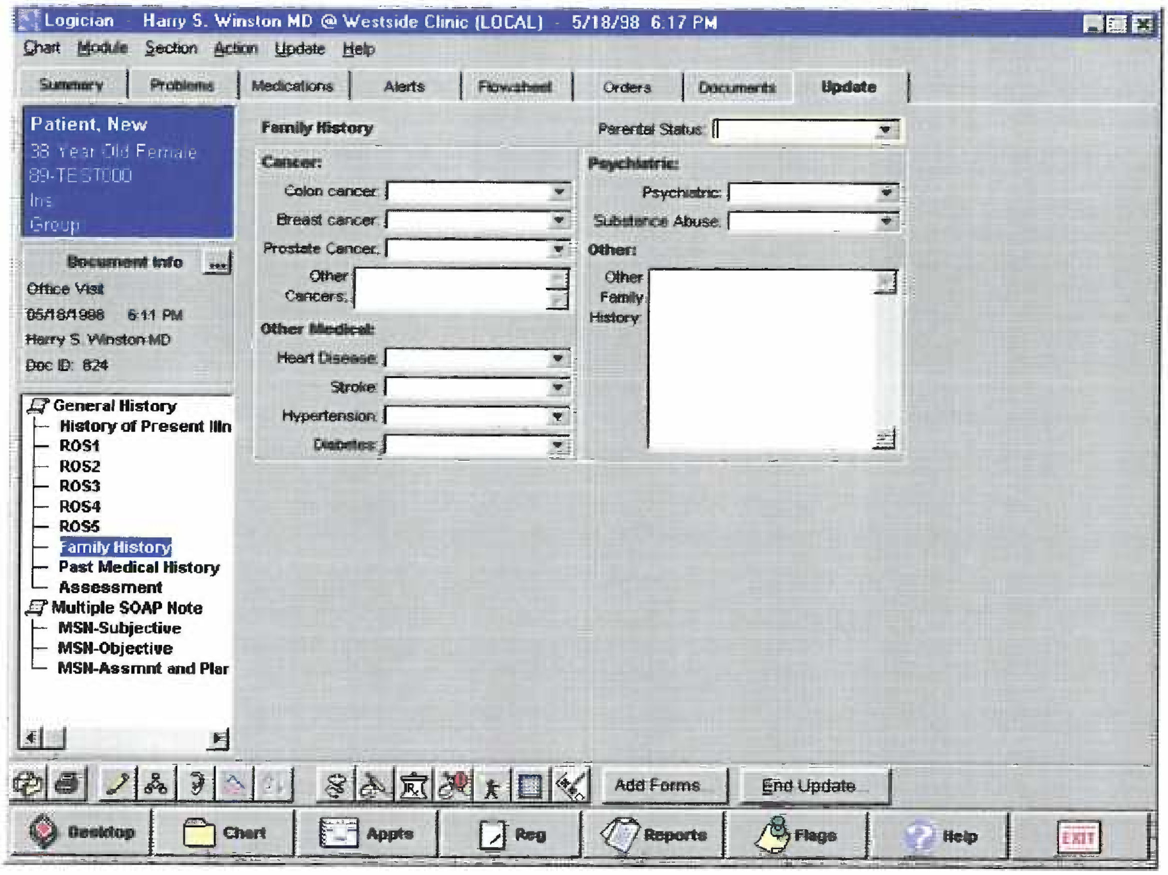

Logician Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6:15 PM
Chart Module Section Action Update Help
Summary Problems Medications Alerts Flowsheet Orders Documents Update
Patient, New
38 Year Old Female
89-TEST000
Ins
Group
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
System
Male GU:
Patient complains of:
dysuria
hematuria
discharge
urinary frequency
urinary hesitancy
nocturia
incontinence
Patient denies:
dysuria
hematuria
discharge
urinary frequency
urinary hesitancy
nocturia
incontinence
Comments:
System
Musculo -skeletal:
Patient complains of:
back pain
joint pain
joint swelling
muscle cramps
muscle weakness
stiffness
arthritis
Patient denies:
back pain
joint pain
joint swelling
muscle cramps
muscle weakness
stiffness
arthritis
Comments:
Add Forms...
End Update...
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT

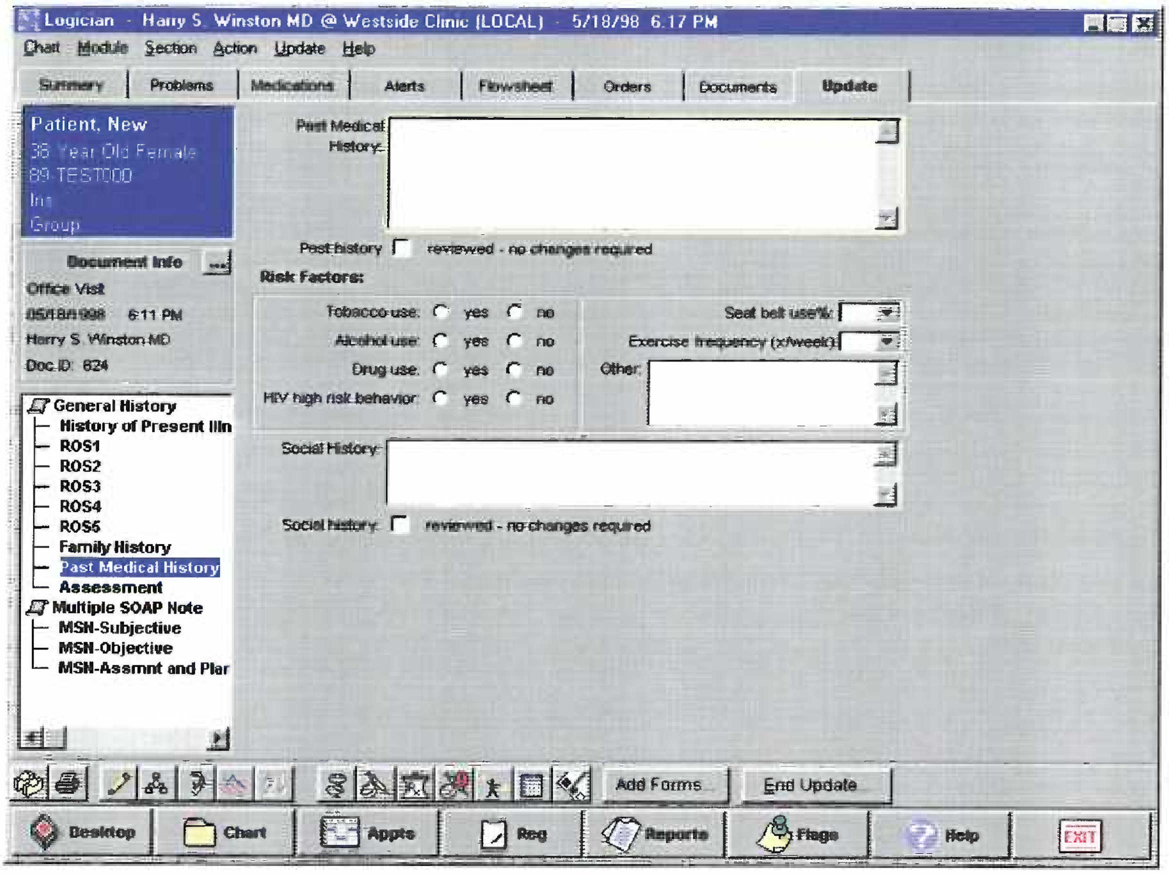

Logician - Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6:16 PM
Chart Module Section Action Update Help
Summary
Problems
Medications
Alerts
Flowsheet
Orders
Documents
Update
Patient, New
38 Year Old Female
89-TEST000
Ins
Group
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
System
Patient complains of:
Patient denies:
Skin:
rash
itching
dryness
suspicious lesions
Comments:
Breast:
lumps
pain
nipple discharge
Neurologic:
transient paralysis
weakness
paresthesias
seizures
syncope
tremors
vertigo
Add Forms...
End Update...
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT

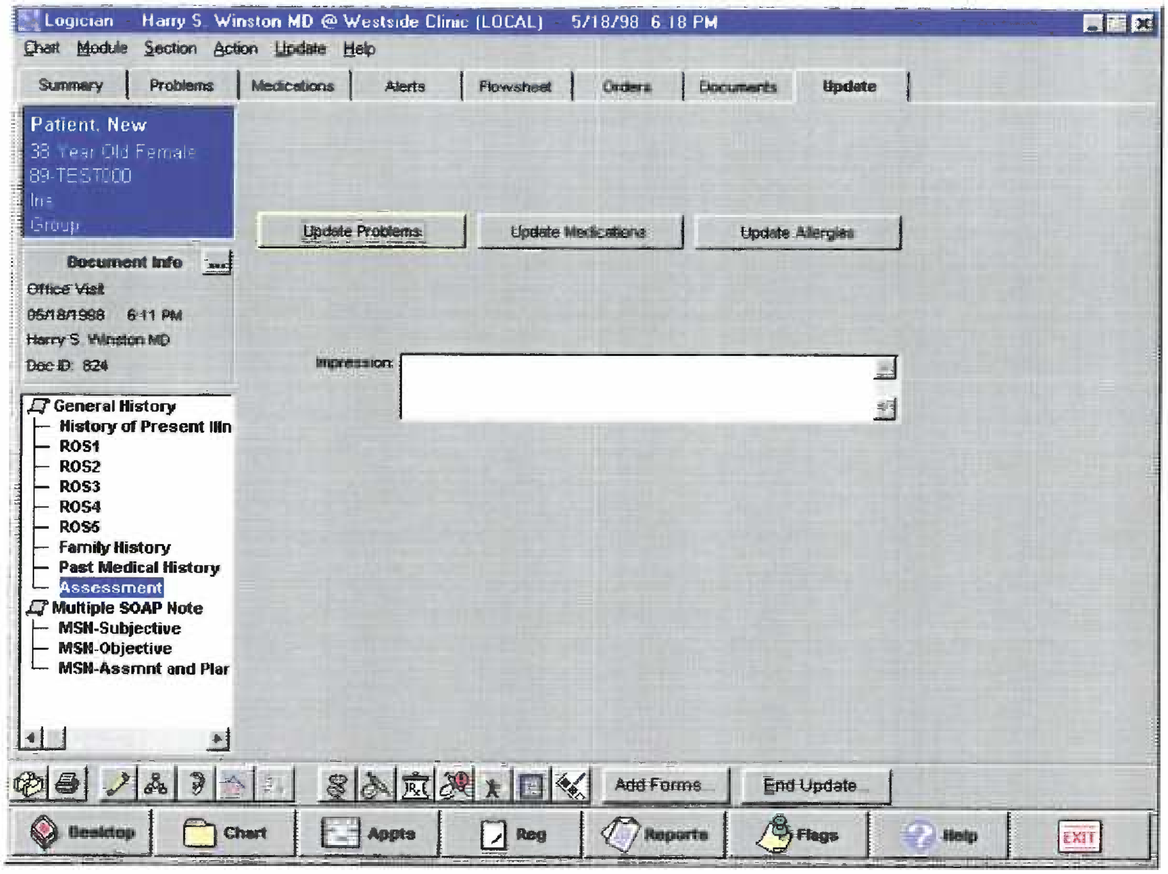


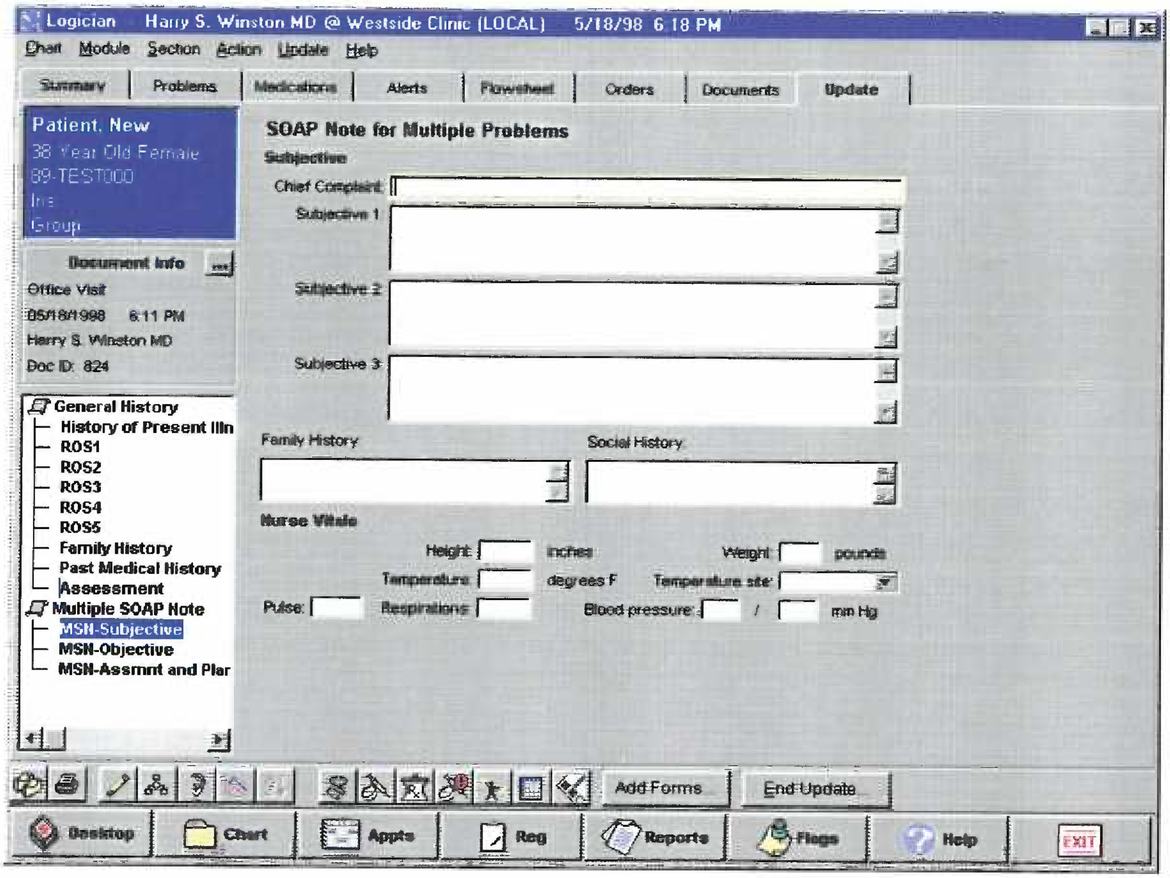
Logician - Harry S. Winston MD @ Westside Clinic (LOCAL) 5/18/98 6:16 PM
Chart Module Section Action Update Help
Summary Problems Medications Alerts Flowsheet Orders Documents Update
Patient, New
38 Year Old Female
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
System
Psychiatric:
Patient complains of:
depression
anxiety
tension
memory loss
mental disturbance
Patient denies:
depression
anxiety
tension
memory loss
mental disturbance
Comments:
System
Endocrine:
Patient complains of:
cold intolerance
heat intolerance
polydipsia
polyphagia
polyuria
weight change
Patient denies:
cold intolerance
heat intolerance
polydipsia
polyphagia
polyuria
weight change
Comments:
System
Hematologic Lymphatic:
Patient complains of:
abnormal bruising
bleeding
lymph nodes
Patient denies:
abnormal bruising
bleeding
lymph nodes
Comments:
System
Allergic Immunologic:
Patient complains of:
urticaria
persistent infections
HIV exposure
Patient denies:
urticaria
persistent infections
HIV exposure
Add Forms
End Update
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT

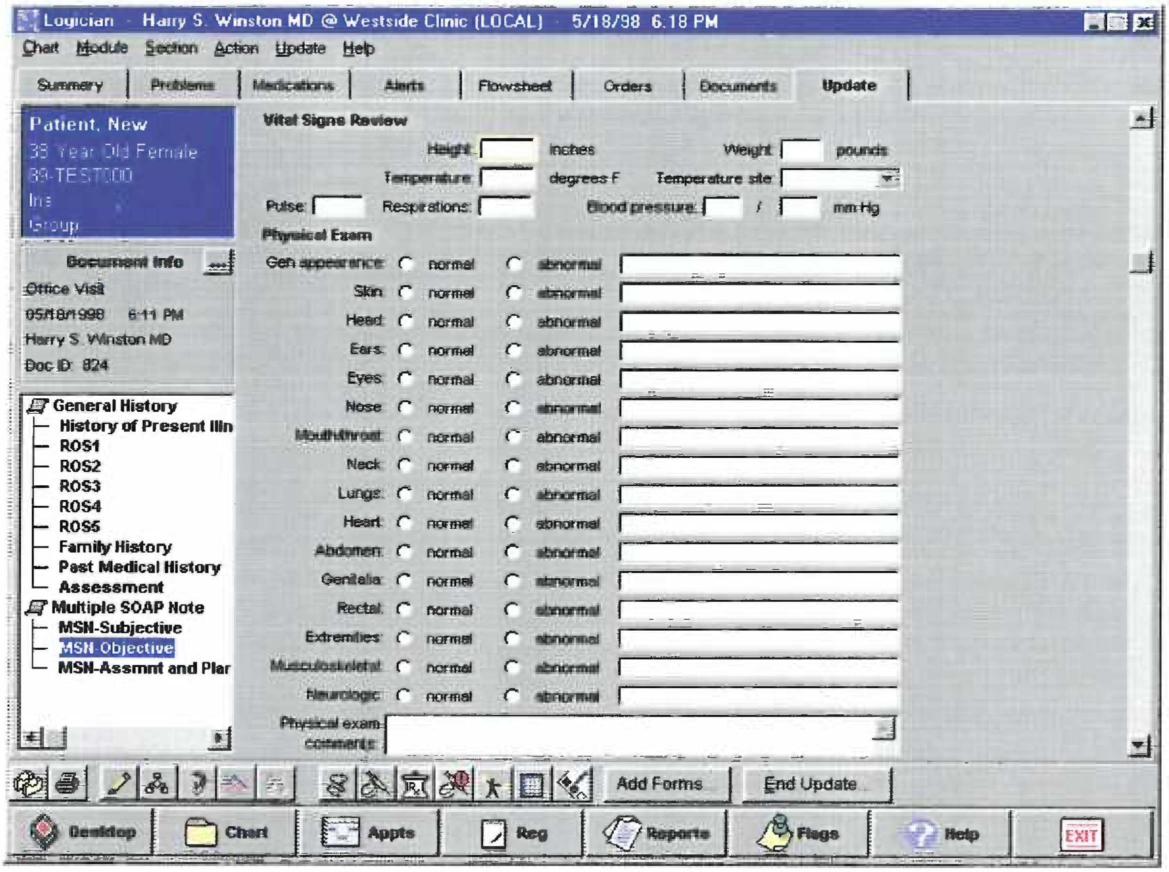

Logician - Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6:16 PM
Chart Module Section Action Update Help
Summary Problems Medications Alerts Flowsheet Orders Documents Update
Patient, New
38 Year Old Female
89-TEST000
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
System
Endocrine:
Patient complains of:
cold intolerance
heat intolerance
polydipsia
polyphagia
polyuria
weight change
Patient denies:
cold intolerance
heat intolerance
polydipsia
polyphagia
polyuria
weight change
Comments
System
Hematologic Lymphatic:
Patient complains of:
abnormal bruising
bleeding
lymph nodes
Patient denies:
abnormal bruising
bleeding
lymph nodes
Comments
System
Allergic Immunologic:
Patient complains of:
urticaria
persistent infections
HIV exposure
Patient denies:
urticaria
persistent infections
HIV exposure
Comments
Add Forms...
End Update...
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT

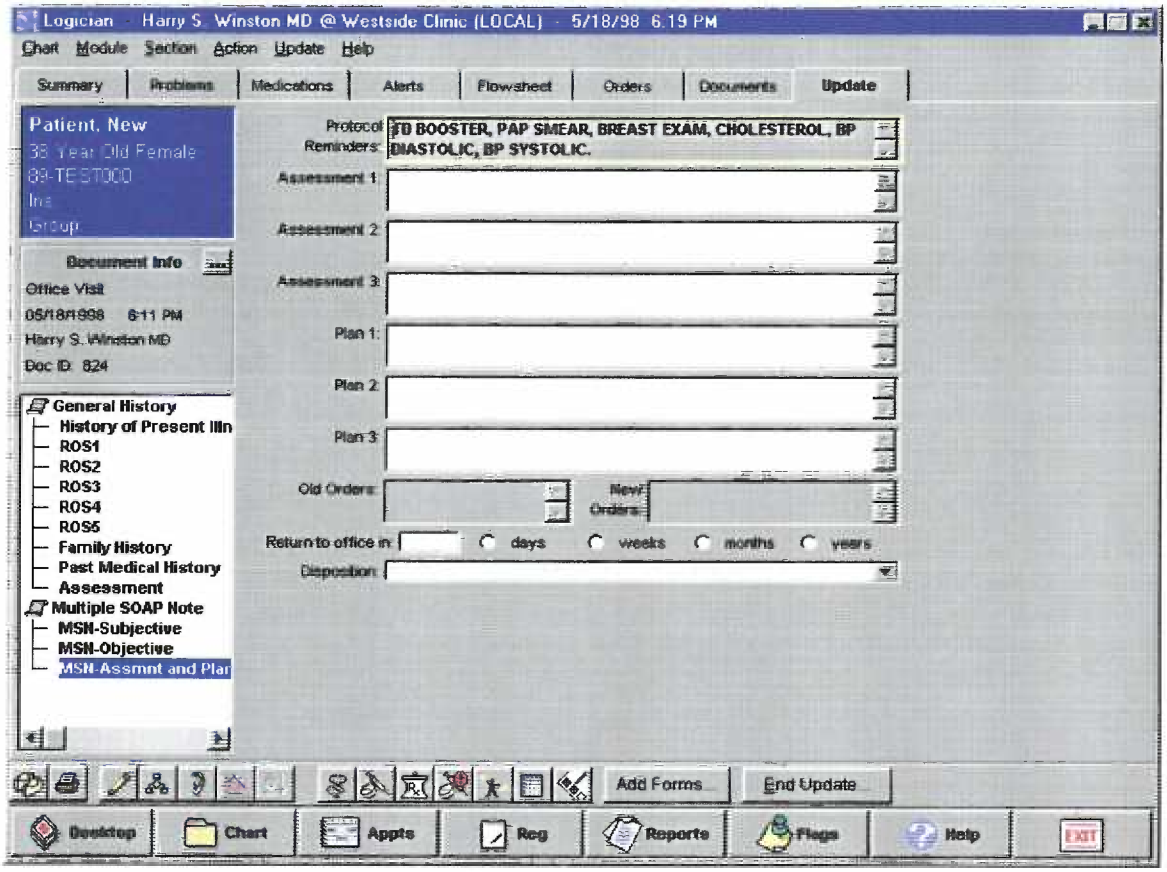

Logician - Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6:17 PM
Chart Module Section Action Update Help
Summary
Problems
Medications
Alerts
Flowsheet
Orders
Documents
Update
Patient, New
38 Year Old Female
89-TEST000
Ins
Group
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
Family History
Parental Status:
Cancer:
Colon cancer:
Breast cancer:
Prostate Cancer:
Other Cancers:
Other Medical:
Heart Disease:
Stroke:
Hypertension:
Diabetes:
Psychiatric:
Psychiatric:
Substance Abuse:
Other:
Other Family History
Add Forms
End Update
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT


Logician - Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6:17 PM
Chart Module Section Action Update Help
Summary Problems Medications Alerts Flowsheet Orders Documents Update
Patient, New
38 Year Old Female
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 624
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
Past Medical History:
Past history reviewed - no changes required
Risk Factors:
Tobacco use: yes no
Alcohol use: yes no
Drug use: yes no
HIV high risk behavior: yes no
Seat belt use%:
Exercise frequency (x/week):
Other:
Social History:
Social history reviewed - no changes required
Add Forms
End Update
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT


Logician Harry S. Winston MD @ Westside Clinic (LOCAL) 5/18/98 6:18 PM
Chart Module Section Action Update Help
Summary
Problems
Medications
Alerts
Flowsheet
Orders
Documents
Update
Patient, New
38 Year Old Female
89-TEST000
Ins
Group
Update Problems
Update Medications
Update Allergies
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
Impression:
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS6
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
Add Forms...
End Update...
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT


Logician Harry S. Winston MD @ Westside Clinic (LOCAL) 5/18/98 6:18 PM
Chart Module Section Action Update Help
Summary
Problems
Medications
Alerts
Flowsheet
Orders
Documents
Update
Patient, New
38 Year Old Female
89-TEST000
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plar
SOAP Note for Multiple Problems
Subjective
Chief Complaint
Subjective 1
Subjective 2
Subjective 3
Family History
Social History
Nurse Vitals
Height
inches
Weight
pounds
Temperature
degrees F
Temperature site
Pulse
Respirations
Blood pressure
mm Hg
Add Forms
End Update
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT


Logician - Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6:18 PM
Chart Module Section Action Update Help
Summary
Problems
Medications
Alerts
Flowsheet
Orders
Documents
Update
Patient, New
38 Year Old Female
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plan
Vital Signs Review
Height
inches
Weight
pounds
Temperature
degrees F
Temperature site
Pulse
Respirations
Blood pressure
mmHg
Physical Exam
Gen appearance
Skin
Head
Ears
Eyes
Nose
Mouth/throat
Neck
Lungs
Heart
Abdomen
Genitalia
Rectal
Extremities
Musculoskeletal
Neurologic
normal
abnormal
Physical exam comments
Add Forms
End Update
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT


Logician - Harry S. Winston MD @ Westside Clinic (LOCAL) - 5/18/98 6:19 PM
Chart Module Section Action Update Help
Summary
Problems
Medications
Alerts
Flowsheet
Orders
Documents
Update
Patient, New
38 Year Old Female
Document Info
Office Visit
05/18/1998 6:11 PM
Harry S. Winston MD
Doc ID: 824
General History
History of Present Illn
ROS1
ROS2
ROS3
ROS4
ROS5
Family History
Past Medical History
Assessment
Multiple SOAP Note
MSN-Subjective
MSN-Objective
MSN-Assmnt and Plan
Protocol Reminders:
TD BOOSTER, PAP SMEAR, BREAST EXAM, CHOLESTEROL, BP DIASTOLIC, BP SYSTOLIC.
Assessment 1:
Assessment 2:
Assessment 3:
Plan 1:
Plan 2:
Plan 3:
Old Orders:
New Orders:
Return to office in:
days
weeks
months
years
Disposition:
Add Forms...
End Update...
Desktop
Chart
Appts
Reg
Reports
Flags
Help
EXIT

Appendix B: Abstracted Data

Compiled data abstraction forms

Patient B 80-82

Patient D 83-84

Patient B - Page 1

PATIENT B	DICTATION				LOGICIAN			
CLINICIAN	2	5	6	8	1	3	4	7
•cc: abdominal pain	-	+	+	+	+	•+	+	+
cc: needs a new PCP	+	+	+	+	-	-	-	-
HPI:								
•abdominal pain	+	+	+	+	+	•+	+	+
intermittent	+	+	+	+	-	+	+	•many times a day
10 years	+	+	+	+	+	+	+	+
•more severe (affecting lifestyle)	+	+	+	+	•+	•+	+	•+
•2/several months	+	+	+	+	+	+	+	•+
epigastric	+	+	+	+	+	+	+	+
LUQ	+		-	+	+	+	+	+
no change with food	+	+	-	+	+	-	+	+
intensified with movement	+	+	+	+	+	+	+	+
no change with metamucil	+	+	+	+	+	+	+	+
no change with ibuprofen	+	-	+	+	+	+	+	-
no pain medications used	+	-	-	-	-	-	-	-
associated with:								
•decreased appetite	+	-	+	+	•Denies anorexia	+	+	•+
• possible weight loss	+	-	+	+	•Denies weight loss	+	+	•+
normal weight 112#	+	-	+	+	-	-	-	-
high fiber diet	+	-	+	+	-	-	+	-
•sometimes constipated	+	+	+	+	•+	+	+	•+
•loose stools with mucous	+	+	+	+	+	+	+	•+
stress-induced	+	+	+	+	-	-	+	+
•no bloody stools	+	+	-	+	•+	+	+	•+
•no melena	+	+	-	+	•+	+	+	•+
prior evaluation:								
exam by gynecologist (implied normal)	+	+	+	+	+	+	+	+
exam by gastroenterologist	+	+	-	+	-	-	+	+
endoscopy performed	-	+	+	+	+	+	+	+

Patient B - Page 2

normal	-	+	+	+	+	+	+	+
colonoscopy performed	-	+	+	+	•barium enema	+	+	+
diverticulae	+	+	+	+	+	+	+	+
CT scan performed	-	+	+	+	+	+	+	+
normal	-	+	+	+	+	+	+	+
ROS:								
sleep problems	+	+	-	-	-	+	+	-
•lack of energy/motivation/lethargic/tired	+	+	-	+	•Denies fatigue, malaise	+	+	•+
no suicidal ideation	+	+	-	+	-	+	+	-
PMH:								
vaginal hysterectomy	+	+	+	+	+	-	+	+
6-7 years ago	+	-	-	+	+	+	+	+
caused by vaginal bleeding	+	-	+	+	+	+	-	+
hypertension	+	+	+	+	+	-	+	+
x 5 years	+	-	-	+	+	+	+	+
controlled with meds/normal now	-	-	-	+	-	+	-	+
Allergies:								
none	+	+	+	+	+	-	+	-
Medications:								
Prinivil	+	+	+	+	+	+	+	+
Premarin	+	+	+	+	+	+	+	+
.. 0.625	+	+	+	+	+	-	+	+
Multiple vitamins (A,D,E B6,B12,CA)	+	+	+	+	-	+	+	+
Family History:								
mother died	+	-	-	-	-	-	+	+
pancreatic cancer	+	+	+	+	+	+	+	+
uncle died bone cancer	+	-	+	+	-	+	+	+
no history colon cancer	+	-	-	-	-	+	+	+
no history early heart disease	+	-	-	-	-	+	-	+
possible FH diabetes	+	+	+	-	-	+	-	-

Patient B - Page 3

SH:								
married	+	+	+	+	+	+	+	+
lives with husband	+	-	+	-	-	+	+	+
retired	+	-	-	-	-	-	+	-
unemployed/housewife	+	+	+	+	+	+	+	-
•former smoker (quit 5 years ago)	+	+	+	+	•+	+	+	•Tob use: no
<1/2 ppd	+	-	+	+	-	-	-	-
•occasional wine	+	+	+	+	•+	•+	•+	•+
•no illegal drugs	+	+	-	+	•+	•+	+	-
Safety Issues:								
•uses seat belts	+	-	+	+	•+	+	•+	•+
no firearms in house	+	-	+	+	-	-	+	-
functional smoke detector in home	+	-	+	+	-	-	-	-
no fire extinguisher in home	+	-	+	+	-	-	+	-
EXTRAS:		she notes she has been feeling depressed			•?2.5 mg tabs	•Pt c/o depression	Readily admits to feeling depressed	•Pt c/o depression, anxiety
		pap smear			•HIV hi risk:no			•denies N,V
		pain is achy						•pain when awake
								•HIV hi risk:no