

Grant Final Report

Grant ID: R18 HS 017060

**Closing the Feedback Loop to Improve
Diagnostic Quality**

Inclusive dates: 09/01/07 - 08/31/11

Principal Investigator:

Eta S. Berner, EdD (Eta Berner Weiss)

Team Members:

Jeffrey Burkhardt, PhD
Mark Cohen, MD
Thomas English, PhD
Shannon Houser, PhD
Marc Krawitz, MBA
Richard Maisiak, PhD
Christa Nevin, MS

Midge Ray, MSN,RN
Anantachai Panjamapirom, PhD
Gordon Schiff, MD
Anne Schmidt, MD
Nicole Shevin, BS
Robert Wears, MD
James Willig, MD

Performing Organization:

University of Alabama at Birmingham

Project Officer:

Kevin J. Chaney, MGS

Submitted to:

The Agency for Healthcare Research and Quality (AHRQ)
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Abstract

Purpose: To develop automated patient follow-up and physician feedback in ambulatory settings.

Scope: This project developed methods to close the “feedback loop” for physicians in ambulatory settings. The main study was carried out in three different sites. A secondary study was done in the Emergency Department (ED).

Methods: Main Study: There were three data collection phases: Follow-up data from patients obtained (1) three weeks after visit via telephone with no physician feedback, (2) one week after visit via telephone with feedback, and (3) one week after visit via interactive voice response (IVR) with feedback. Analyses included problem resolution, medication adherence, patient satisfaction, and physician satisfaction, and cost impact. Secondary study: A method for assessing concordance between Emergency Room diagnoses and hospital discharge diagnoses was developed. Formative evaluation of methods of providing feedback to residents was conducted.

Results: Main Study: For phases 1 and 2 approximately 15% of problems were unresolved. For the IVR phase, fewer patients were reached and more of those who were reached had unresolved problems. Patient satisfaction was high during all phases, and was higher in those who received follow-up calls. Physicians who viewed the feedback found it helpful. Cost analyses showed that if a follow-up system were implemented routinely the expense could be offset by increased revenue of the return visits, with the potential to improve the quality of care and avert higher costs of hospitalizations. Secondary Study: Overall dissonance rates between ED and discharge diagnoses was approximately 10%.

Key Words: Interactive Voice Response, Feedback, Ambulatory Care

The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services of a particular drug, device, test, treatment, or other clinical service.

Final Report

Purpose

One of the purposes of the EQM RFA was to “investigate novel methods...for the development of Health IT to assist clinicians, practices and systems to measure the quality and safety of care in ambulatory settings.” The particular focus of the RFA that we targeted in this project was to use health IT to deliver timely feedback for improvement to physicians in ambulatory care settings. Systematic feedback to physicians of patient outcomes in ambulatory settings is rare. As stated in the Institute of Medicine (IOM) report, Preventing Medication Errors (page 203), “In the ambulatory setting, medication monitoring, particularly for ADEs is virtually non-existent.”¹ Both Schiff and Wears have commented on the importance, but unfortunately frequent absence of, timely follow-up of patients in ambulatory settings.^{2,3} The usual process in ambulatory care is an open loop, where the physician assesses the patient and rarely learns whether the assessment and treatment had the desired effect. The absence of such feedback can lead to incorrect diagnoses failing to be detected, medication adverse events, and systematic problems with medication adherence not getting addressed. This project was aimed at developing efficient mechanisms to close this feedback loop, to deploy them in routine ambulatory care settings, and to assess the impact on safety and quality. Our specific aims were to:

1. Develop a system within three different ambulatory electronic medical record systems in three different types of ambulatory settings that includes (1) pro-active follow-up of patients’ response to treatment (including medication adherence and adverse events) and (2) feedback to health care providers;
2. Assess the impact of automating the follow-up/feedback process. Impact was measured in terms of:
 - a. Diagnostic quality (problem resolution)
 - b. Patient satisfaction with their clinical care
 - c. Physician satisfaction with the feedback process
 - d. Resource Utilization; and,
3. Develop and evaluate an automated system for feedback to Emergency Medicine physicians of the concordance between their diagnoses and patients’ final diagnostic outcomes.

Scope

Background and Context

Our study was designed to address the problem of follow-up of patients in ambulatory settings and feedback to the patient's physician. We focused on acute care problems on the assumption that patients with chronic illness were more likely to see a physician regularly, but acute care visits are more likely to be episodic with less potential for feedback to the physician.

Study Sites

The intervention involved developing automated processes for proactive patient follow-up and ongoing feedback to physicians. The intervention was carried out at four varied sites—three ambulatory clinic settings and one Emergency Department setting. The clinics include a Family Medicine Clinic at the University of Alabama at Birmingham's (UAB) Huntsville, Alabama campus (henceforth referred to as Family Medicine), UAB's HIV clinic in Birmingham (HIV), and the comprehensive evaluation center of United Cerebral Palsy of Greater Birmingham (UCP). These sites all have different electronic medical records (Allscripts, an internally developed system, and Office VistA, respectively). Both Allscripts and Office VistA are certified EHRs. Because we were using sites who already had implemented EHRs, we did not assess the impact of our project in low resource settings. Our focus was on providing feedback to physicians and patient access to their health information was not within that scope.

This study included the following priority populations: low income, minority, inner city and rural, women, elderly, and those with special health care needs (HIV and Cerebral Palsy). Although many of the clinic patients had chronic conditions, for this study, acute care visits were targeted. This meant that many of the IOM priority areas were not directly targeted, although some of our patients may have had acute exacerbations of their chronic conditions and were included. The main IOM priority area that we did target was medication management since we were assessing medication adherence. The involved professionals included residents and fellows, attending physicians, nurses, and allied health professionals in both primary care and specialty practice. The focus in all three of the clinic sites was on follow-up of, and feedback on, patients who are treated on an outpatient basis.

The Emergency Medicine site was Shands Hospital, an affiliate of the University of Florida, in Jacksonville, Florida. The focus there was on feedback to the Emergency Room residents on the diagnoses of patients admitted to the hospital. Below we refer to the main and secondary studies. The main study was carried out at ambulatory clinics at the following three sites:

1. HIV
2. Family Medicine
3. UCP

The secondary study was carried out at the following site:

- Shands Hospital Emergency Department (ED)

The main study was approved by the Institutional Review Board (IRB) at the University of Alabama at Birmingham and the secondary study was approved by the IRB at the University of Florida. The methods and results for each study are described separately.

Methods

Main Study: Specific Aim 1—Three Phases of Data Collection and Feedback

Phase 1: Baseline Data Collection with No Automation and No Feedback. In this phase patient outcome data, including response to treatment, medication adherence and adverse events, were collected three weeks after the initial patient visit. This time interval was planned with the assumption that most problems would have resolved, but we also assessed what happened during the three week interval. Data included patient report of their current health status, problems with any of the medications that were prescribed during this visit, other actions that they took in regard to the problem, and general medication adherence. This assessment was carried out at all three sites by a study staff person who worked at the clinic. At the UCP clinic, the assessment was done by an interview and at the other two sites via telephone. Any patients who still had problems were immediately transferred to the usual channels of the clinic for handling patient phone calls. We adopted an opt-in consenting procedure. Patients were given information about the follow-up plan while they were in the clinic and we obtained contact information for the phone call from them at that time. To authenticate patients so that we could be sure we were talking to the actual patient, not just anyone who answered the telephone, we suggested patients give us the last four digits of their social security number or another number that they would easily remember and we used that as an authentication code. Verbal consent was obtained at the time of the phone call.

Phase 2: One Week Assessment with Partial Automation. Although initially this assessment was conceived to be done at varying intervals depending on the reason for visit, this plan proved to be infeasible. Based on clinician and patient input we determined that the one-week interval would be soon enough after the visit to identify problems before they became very severe, soon enough for patients to recall the information, yet enough time to allow therapies to work. The questions were the same as the baseline data collection and this phase was carried out at all sites in a manner similar to the baseline assessment. However, during this phase a feedback report to the physician was initiated. In the Birmingham sites this feedback was automated so that when physicians logged onto the EMR they had the option of viewing the feedback report. In the Huntsville site, messaging features of the EMR were used for the report. To conduct the follow-up and provide the feedback we developed a system we refer to as CDNA (Clinical Documentation and Notification Application). CDNA is currently being reviewed by the UAB Research Foundation for its patent potential.

CDNA was used to schedule follow-up phone calls to participating patients, extract patient and visit data from each site's EMR system, collect the data from the follow-up calls to generate a feedback report to providers that was displayed when providers logged onto their EMR. CDNA can be interfaced with a human or with an interactive voice response (IVR) system.

In Phase 2, CDNA was used by the study personnel at both the HIV clinic and the UCP site. When study personnel used it, they were able to see which patients were scheduled for a given day's interview and what their diagnoses and medications were. They could enter the results of the telephone interview they conducted into a form that added the data to CDNA.

Phase 3: One Week with Full Automation. This phase was similar to Phase 2, but instead of the clinic staff making the phone calls, an IVR system made the phone calls. In addition, it was determined that the IVR approach would not be appropriate for the UCP site, so this phase was only done in the HIV and the Family Medicine clinics. The data collection questions for the IVR were similar to those for the other two phases, with slight modifications for use with the IVR. CDNA was used for scheduling the calls, storing the data and preparing the report. This automation was accomplished by (1) CDNA extracting data from the EHR and storing it in the CDNA database (2) the IVR using the data to make the phone calls, (3) the IVR transmitting the patient's responses back into the CDNA database and (4) CDNA transmitting a report back to the EHRs in the two sites. In the HIV clinic this process occurred entirely automatically. In the Family Medicine site, in part because of the different EHR system and in part because of different institutional policies, there were some manual interfaces in the process. All data were stored and transmitted in a secure and HIPAA-compliant manner that met the stringent security policies at both sites.

Main Study: Specific Aim 2—Additional Data Collection and Analysis

For each phase we assessed the percentage of patients whose problems were still unresolved, or who had problems with their medications. This data on outcomes included an ordinal scale (much better, better, same, worse, much worse) as well as descriptive data in some cases. In addition, our plans were to compare the problem resolution of patients three weeks after the visit when the automated data collection and feedback process was in place during Phase 3 with the baseline data collected three weeks after the visit with no feedback during Phase 1. That analysis could not be reliably done for several reasons: (1) The problems of almost 90% of the patients at baseline, and even at one week, were resolved and (2) many of the patients in phase 3 with the IVR did not complete the interviews. For these reasons, outcome assessment focused on percentage of patients who were not improved at each phase.

In addition to the data collection from patients after specific ambulatory visits, we also collected patient reported experience and satisfaction with their care from a sample of all patients seen in the clinics, some of whom also received the follow-up calls. Although we did not specifically assess how these data were used in the clinics, to assure that the data were more likely to be used, we adapted some of the questionnaires that they were already using. These questions addressed some similar areas to the Consumer Assessment of Healthcare Providers and Systems (CAHPS) questions including clinic communication, overall satisfaction, recommendations for improvement, and assessment of the concept of follow-up calls and for those who actually received follow-up calls, their opinion of the calls themselves. We asked the patients about their perceptions of the quality of care but did not provide them with other reports

on their providers' quality. These data were collected anonymously after each phase of the study ended at two of the sites. The data were collected via an interview at the United Cerebral Palsy site, but because of the difference in data collection methods and because so few patients were interviewed they were not included in the analyses. Because we did not use the actual CAHPS questionnaires we did not provide any data to the National CAHPS Benchmarking Database.

Data collection from physicians assessing their reaction to the feedback was collected via a questionnaire after the third phase of the study. We assessed the percentage of providers who used the system to evaluate their patients' experience, and also obtained an assessment of its usefulness, specifically whether physicians reviewed the feedback reports, whether they found them useful and whether they would like to continue to receive them.

We assessed several aspects of costs. One aspect was the costs of actually doing the study. We collected data on the costs of recruiting subjects for the study as well as conducting the phone interviews. The time expended at each of the sites on recruitment and enrollment and follow-up interviews was collected and converted into costs. Each site submitted the time required for each step of the recruiting and enrollment process. Some tasks were able to be measured precisely, while oftentimes, the number of minutes required at each task was estimated. Using salaries for the people performing each of the tasks, the minutes used were converted to dollars spent, thus giving an estimate of the cost of implementing the recruiting and enrollment process at each site.

The second aspect of cost analysis was examining the cost impact of implementing the system based on an estimate of the patients who could benefit from the follow-up. The rationale and analysis are described in more detail in the results.

Medication adherence was assessed by modifying (and validating) the Morisky medication adherence scale that has been used for hypertension medication adherence.⁴ We modified a recent update of the scale to be able to be used for more general adherence.⁵ These data were collected on the follow-up call.

Secondary Study

Procedures for Retrieval and Matching Diagnoses. We convened a work group of the relevant organizational and technical experts to accomplish this task, which was complicated by the administrative and financial organization of work in the health center. Data on diagnoses and relevant covariables are distributed over two different information systems from two different vendors, one for physician billing and coding, and another for hospital billing and coding; Emergency Department (ED) diagnoses are represented in the physician billing database, while the final discharge diagnoses were maintained in the hospital database. The procedure for identifying appropriate patient records was as follows:

1. At a specified interval (e.g., monthly), the hospital billing database (McKesson) is scanned for discharges in the preceding time period, and a filter applied that selects only patients who were admitted to the hospital from the ED. This resulted in a text file containing patient identifiers and final discharge diagnoses as ICD-9 codes.
2. This file was then used to match records in the physician billing dataset (IDX) for the same period; the matching process was based on unit number, date of birth, SSN, and date of service +/- 24 hours. Records in the physician database were extracted as text

files containing patient identifying information, ED physician identifying information and the ED ICD-9 diagnoses codes.

3. The ED and final discharge data sets were then merged based on patient identifiers and date of service to produce a single data set.

This process was pilot-tested and slightly revised for efficiency. The primary adjustment made was, instead of identifying relevant patients at the point of admission (when final diagnosis information may not be available), they were instead identified at the time of discharge and the admission level information matched at that time.

The revised process was piloted successfully and encountered only a few problems. There were sporadic cases where more than one physician billing record was matched to the same ED visit and a small proportion of unmatched records. We were unable to identify the source of these discrepancies and since the total number of discrepant matches was small, we treated those observations as missing.

Procedures for Adjudicating ED and Final Diagnoses. Four physicians participated in the adjudication procedure; they were trained and experienced in emergency medicine and internal medicine / critical care or both specialties. To determine the degree of agreement on the diagnoses we used implicit criteria and a process of training and mutual discussion to develop a reasonably consistent set of judgments. After multiple joint coding sessions in which the entire group reviewed the data records and used access to the chart to resolve items of discussion, we were able to codify our criteria as follows. We used the terms ‘consonance’ and ‘dissonance’ to try to reflect a sense of value-neutrality.

1. The fundamental judgment to be made was agreement or disagreement among the diagnoses on the two lists based on clinical relevance and importance.
 - a. If a diagnosis were viewed as possibly dissonant, the feeling that “no one could or should possibly have known that” was considered irrelevant, and the results coded as dissonant.
 - b. Similarly, technical disagreement – ‘acute bronchitis’ vs ‘pneumonia’ vs ‘pulmonary infiltrate’ would usually be considered consonant
2. The final diagnosis list was expected to be more nearly comprehensive, in the sense of listing chronic conditions, than the ED list, which is expected to be problem focused. Thus, the omission of stable chronic conditions on the ED diagnosis list would normally not be considered dissonant.
3. We did not assume either list was privileged or presumed to be correct, *a priori*. The more nearly correct list should be apparent from the lists, or by referring to the chart etc. Thus it may be possible that both sources could be sources of dissonance (see below).

4. If indeterminate on initial review, the reviewer reviewed the scanned chart for clarification, using the same principles. If still indeterminate, final adjudication was made via a second or group review of the record.
5. If the diagnosis lists were considered consonant, an additional judgment was required to be sure that one or the other was not overly broad, e.g., ‘malaise’ or ‘fever’ vs ‘sepsis’.
6. If dissonant, an additional judgment was required to identify the source of the dissonance. Typically this was either the ED or the inpatient list, but in principle it could be both, or could occasionally be attributed to a coding error.
7. We explicitly made no attempt to assess the severity or clinical impact of any discrepancies in diagnosis.

Feedback to Residents. We piloted a number of modalities for providing feedback. These methods and the reasons for changing approaches are described in the results.

Results

Main Study

Table 1 shows the number of patients involved in each phase of the study, excluding the patient satisfaction data which are reported separately.

Table 1. Number of patients in each study phase

	Baseline (3 Week follow-up)	Partial automation (1 Week Follow-up)	IVR (1 Week/2 sites)	Total
Called	142	352	122	616
Reached	113 (80%)	203 (60%)	82 (67%)	398
Consented	111 (98% of those reached)	197 (97% of those reached)	45 (55% of those reached)	353
Completed	109 (98% of those consented)	197 (100% of those consented)	38 (84% of those consented)	344

Problem Resolution. At baseline, three weeks after the visit, approximately 12% of the patients who consented to the interview reported no improvement or were worse, with approximately half being worse. At one week with the clinic staff calling, 15% showed no improvement or were worse, but most of those were no improvement. At the one week phone call done via the IVR, 36% of those consented were the same or worse, and again most of those reported no improvement. It should be remembered, however, that for the baseline and partial automation calls, almost all of the patients reached consented to participate, but for the IVR calls

only 55% consented to the interview and more of those in the IVR condition did not complete the entire interview. Another study of outbound health care follow-up with IVR also found high rates of abandoned calls.⁶ It is difficult to know if this is typical in other industries because, according to a review of call centers (human and automated) by Gans and colleagues, most of the research has been done on inbound calls, where the client calls into the system, not outbound calls such as we implemented.⁷ It may be reasonable to assume that those who did not consent, or who hung up during the phone call, most likely did not have anything to report and were less constrained about hanging up on a “computer” than the live interviewer who called in the other phases. If we use as the denominators the number reached rather than the number consented, the percentages who are the same or worse are much closer (12%, 14% 18%).

Of those who reported being better during any of the three data collection phases, over 80% had not contacted anyone, which is not surprising. However, even after three weeks, 23% of those who were not improved or were worse also did not contact anyone.

At one week, as might be expected, because most patients who reported they were not improved were also not worse, between 68% and 86% of them had not contacted anyone. When contact was made, overall, the most frequent actions were to call and/or make an appointment with their own doctor. These data show that a surprising number of patients do not take action when they have not improved, even after a considerable period of time. In addition approximately 6% wound up seeing another physician or going to the Emergency Room.

During the full automation phase, patients who were not better were followed up two weeks later (three weeks after the initial visit) via IVR to see if the follow-up and feedback made a difference in problem resolution. Unfortunately only five patients responded to the survey. Of those five patients, four of the five were better and only one was worse. Thus we see two sets of patients that could benefit from follow-up calls one week after their visit: (1) those who still were worse or saw no change longer than one week after the visit, so they called the provider at that time, and (2) those who were worse or saw no change at one week and did not take any action on their own.

Medication Adherence. Patients who were prescribed medicines were asked if they filled and took the medications as prescribed. Almost all of the patients at each of the time periods reported that they did take the medicine as prescribed, although a few patients each time had some problems with the medications (did not seem to be working, had reactions, reported that they were allergic) or reported that new symptoms arose. These numbers were very small, indicating that there were not extensive problems with medications or previously undiscovered symptoms. In almost all of these instances the patients stopped taking the medication and/or discussed it with the clinic staff. These results are congruent with the medication compliance reports from those patients who took medicines on a regular basis, which indicated high rates of self-reported compliance.

Patient Satisfaction. A total of 527 patients responded to the patient satisfaction survey after baseline, 306 after the partial automation, and 243 after the IVR implementation. There were no significant differences in overall satisfaction with care among any of the three surveys and the two clinics were similar to each other. Patients were very receptive to the idea of follow-up (88% and 89% on the two surveys where the question was asked said it was a good idea to follow-up after visits). In addition, those who said they received a phone call were pleased with the call. The patients who said they received the phone call from either the clinic staff or the

IVR also had higher overall satisfaction than those who did not receive a call, and the difference in satisfaction between those receiving the call and those who did not receive a call was statistically significant for those receiving the human phone call. A manuscript describing additional details is in preparation.

Physician Satisfaction. A total of 27 physicians responded to the physician satisfaction survey for an overall response rate of 63% (see Table 2).

Table 2. Physician satisfaction response rate

	Family Medicine	HIV Clinic	Total
Number of physicians receiving feedback reports	36	7	43
Number of respondents to Satisfaction Survey	20	7	27
Response rate	56%	100%	63%

Sixty-three percent (63%) of the respondents reported that they would like to continue to get the reports as a part of routine care.

In the Family Medicine clinic the residents logged onto the EMR a minimum of several times a week, with most using it daily. At the HIV clinic, since this was the “sick call” clinic the attendings were not there every day, and use was much less frequent. Over half of the attending logged on just once a week.

Of those who responded, 78% reported that they did look at the feedback reports, but they did this less frequently than the frequency with which they used the EMR. In other words, they reviewed the reports, but not each time they logged on. The majority (76%) of the respondents who reviewed the reports reported that the reports spurred them to do additional follow-up in the patients’ charts. Interestingly, the residents were the most variable in whether they followed up the report with additional chart review. Six residents who reviewed the report did not do additional chart review on any patients, but another six residents did further follow-up in the charts of all patients whose reports they reviewed. Seven residents and three attending physicians at the HIV clinic reported that they tend to review the charts for patients who reported being worse. Five respondents did not answer these questions. A manuscript describing these data in more detail is in preparation.

Resource Utilization. The cost of recruiting and enrolling patients for the different sites varied widely. Differences were due to the different processes involved in finding, enrolling, and interviewing the patient, as well as the ability of the patients to comprehend instructions and interviews. The main reason for the discrepancies (range of approximately \$3.00/patient to almost \$60/patient) depended on how active study staff were in their recruitment efforts. If the follow-up process were simply a routine part of clinic procedures and all patients were followed up, there would be no costs of recruitment. Costs of the human phone calls were more consistent across sites and times, averaging approximately \$3.50 per call, not including the actual phone charges.

We can also look at the impact of instituting this type of follow-up process. At baseline 26% of all patients called had contacted a healthcare provider sometime during the three weeks following their appointment. In the one week non-automated follow-up, 12%_called a healthcare

provider during the one week following their appointment. Assuming that patients were similar in both the baseline and non-automated live human call groups, the difference between the two, (12%), estimates the percentage the patients in the baseline group who called a healthcare provider during the second or third week after their appointment.

Of the total patients called with the IVR one week after their appointment almost 90% had not called a healthcare provider during that one-week period, and of those taking no action, almost two thirds of them reported that they were better at the time of the call. Since there are factors unique to IVR calls, patients may respond to those calls differently than to live human calls. Of all the calls placed by the IVR system, approximately 68% were not completed. This did not occur in the live human calls. However, of those calls that were answered, the percentage of patients who had called a healthcare provider during the first week was similar to that for the live human calls.

From a cost perspective, we can show how much it would cost the provider to make sure that the patients who could benefit are contacted, so that no one falls through the cracks. From our data we estimate that about 10-15% of patients would fall into this category.

Conservatively, then, if a clinic saw 1000 patients per month, and a live follow-up call was made to each of those patients, at a cost of \$3.50 per call (the cost per call in the baseline group), clinic expenses would increase by \$3500 per month. Using baseline numbers, 15%, or 150 patients could benefit from the follow-up call, so on average, the cost to the clinic would be \$23.33 per patient that needed to talk to a provider. If only those who were worse came back for another visit, about 3.7% (two patients who had taken no action, and two patients who called a provider in the second or third week in the baseline group), there would be an additional 37 visits per month. The national average reimbursement for a Medicare level three, established outpatient visit is \$68.97, or an additional \$2552. If a small portion of those who were in the no change category also came for a visit, then the cost of the follow-up calls could be paid through increased patient revenue. We have data on the patients who returned to their primary care provider or who sought out other providers, were sent to the Emergency Room or were hospitalized. We plan additional analyses on the data from this study to determine the potential healthcare costs that timely follow-up could avert.

The calculation of the resource estimates is preliminary at this time and there is much that we cannot tell from our data. We cannot definitively say that implementing the IVR system would save money by catching problems early and preventing expensive medical tests/procedures in the future, although the data are suggestive that approximately 10-15% of ambulatory care patients could benefit from follow-up and being rerouted back into the system. The actual costs of a practice setting up the IVR procedure were not adequately captured with a grant-funded project and we cannot provide specifics on the impact of the additional workload on ancillary staff.

Secondary Study

Retrieving and Matching Diagnoses. Although we still have had occasional discrepancies in matching cases from the hospital IT system and the physicians' IT system, the procedure has worked sufficiently well to this point. Because of the pending comprehensive EHR implementation, further work on refining the system has been suspended but will resume once new system is in use.³

Adjudication of Diagnostic Concordance/Dissonance. Using our adjudication procedure and beginning our coding as a group, we generally experienced 90% agreement with kappas in the range of 0.5 in the initial assessment of diagnostic consonance / dissonance. This level of agreement is generally considered in the moderate range or “fair to good.”⁸ The most important findings are:

1. Overall rate of dissonance was around 10% of cases (7.7 – 13.8). The Emergency Department was the source of the discrepancy in two-thirds of the cases, the remainder equally divided between inpatient services and coding errors.
2. There was no evidence for association of dissonance with acuity (as measured by triage class), admitting service or specialty, admission diagnosis, age, race, or gender.
3. There was no evidence for association with ED length of stay, boarding time, or hospital length of stay.
4. There was no evidence of association with resident or attending physician. Analysis of ‘outliers’ showed no more than expected given a random (Poisson) distribution of dissonances.

A manuscript further describing these findings is in preparation.

Formative Assessment of Methods to Provide Feedback to Residents. We have had a great deal of difficulty in finding an efficient and effective way to provide diagnostic feedback on a routine basis. Provision of summary information only [e.g., an individual’s dissonance rate compared to ED as a whole, or to all providers (unidentified)] had two significant problems. It was prohibitively labor-intensive if comprehensive, and burdensomely labor-intensive when random subsets of cases were reviewed. In addition, feedback from residents suggested summary information was too vague—they wanted specifics about which cases were being flagged, and for what reasons.

The logical response to that problem was to include the diagnosis lists and record numbers of dissonant cases to individual providers. This had the same problem of being too labor intensive. We tried to mitigate that issue by using random subsets of cases, but the review burden was still high. In addition, out of concern for HIPAA violations we converted to a computer-based process.

Because the preceding analysis showed discrepancies to be distributed rather evenly across specialties and acuities, the labor burden could not be easily reduced by focusing on an easily identifiable subset of cases bearing the majority of diagnostic problems.

Review of the diagnosis lists made it clear that straightforward syntactical matching would not be successful. We could not explore the potential for some form of natural language processing functionality to aid in this process as we have no current access to such resources.

We attempted a straightforward, paper feedback process, based either on a comprehensive or random subset of cases (the latter to avoid flooding the provider with so many cases they would not bother to look at them), encountered some of the same HIPAA issues noted previously and this procedure was stopped shortly after we piloted it.

Future Work. We have worked with the capabilities of our graduate education management system and believe we are ready to implement an entirely computer-based, secure, feedback system. This will take the feeds from information services, group them by provider, convert the diagnosis lists and record numbers to pdf files for each provider, and attach those pdf's to the resident's educational record file in the system. The feedback would be password protected, access limited to the resident and the residency director, and access could be monitored. While this process is only semi-automated at present, we believe we can develop scripting that will allow staff to execute the procedure simply when each new dataset is delivered. This system will also allow us to conduct a longitudinal analysis to see if individual's diagnostic consonance rates improve over time.

One remaining issue that we hope to have clarified soon is whether these files will be considered subject to the 'research data requirement' and thus have to be relocated to the secure server, or perhaps encrypted. Our institution has demanded that any research data with PHI be put on a central server. We are hopeful that we can be allowed to view them as required educational feedback and thus obviate that issue.

Discussion

We achieved most of the aims with which we began this study. Most importantly we developed an automated process for feedback and follow-up that was able to work with multiple electronic health record systems, including home-grown and certified EHRs. We also developed and assessed the psychometric properties of a general scale for medication compliance and we were able to show that both patients and physicians were pleased with the "closed loop." We have data showing that although most patients in ambulatory care settings appear to recover as expected, for the minority that do not, there are both clinical and resource implications. If follow-up is done by a human within the clinic, the estimated increase in costs to the provider is likely to be offset by patient care revenue that the early follow-up generates, and if the overall costs are considered, a return phone call or visit to the primary care provider that averts a later emergency room visit and/or hospitalization is certainly likely to be less expensive. The cost of the IVR system (versus human calls) is considerably less than human calls, once it is established. With the IVR system, not all patients were reached, but if the same numbers of problems were identified, and patients came in for appointments, the cost of the calls would be minimal, but the potential to help those patients who were no better would be increased.

In our secondary study, we developed a method to estimate concordance of "downstream diagnoses" with ED diagnoses and have learned lessons on what is needed to provide effective feedback to ED residents.

There were a number of challenges that arose during this study. In terms of the main study, ideally we would have had variable intervals of follow-up depending on the patients' diagnoses, but that proved to be infeasible. There were challenges in terms of both technology and policies in getting our feedback mechanisms developed. For instance, at first we thought that we would directly integrate the feedback reports into the EMR, but that proved not to be feasible. The solution we deployed, however, actually made the process more generalizable, since the reports were able to be accessed without altering the systems' source codes. This process could be fairly easily used by commercial providers of EHRs since it does not depend on a particular EHR technology.

In addition, although the UAB Data Security Officer approved our technical architecture for the automated extracting of data and transmitting them to and from the IVR, the Huntsville IT personnel were less comfortable with automated extraction and transmission to a system at UAB, which necessitated a somewhat different process. We were able to address a variety of other challenges (described below) including developing a system for authenticating patients that satisfied the patients, which is important under any circumstances but especially with an HIV population.

Another challenge was the actual recruitment of patients. Because we were doing this study as research not as routine care, and in part because of IRB recommendations, we chose to consent patients rather than targeting all patients in the clinics, and we only called those who agreed to be contacted. This limited our sample size. Luckily we were able to do a verbal consent on the phone at the time of the interview which was easier than a written consent would have been. Although our sample sizes were not large, we found that for the first two phases almost all of those who agreed to participate did complete the survey and even in the IVR call a large proportion completed it. In another study using IVR, researchers have found that the majority of patients do not answer the calls and if they do answer, a high proportion do not complete the calls.⁶ This challenge may be a limitation of IVR studies and although we had higher proportions of those consenting actually completing the study, our sample sizes were affected.

The diversity of the population of patients we included also posed some challenges. Although the United Cerebral Palsy site was enthusiastic about participation, that population turned out to be a challenge for clinic staff to call and we resorted to in-person interviews and did not attempt the IVR calls. The HIV patients were also a vulnerable population. One of the issues that required a substantial time commitment early on was developing our protocol to have a method of authenticating the patients and also a method of routing them back to the clinic if problems were discovered on the phone call. We felt that this was a safety issue and were not comfortable just collecting the data for our study. In particular, we knew when the full automation was developed and clinic staff were not involved we wanted to assure that patients would get proper attention.

A limitation of our study was that we relied on self-report. While patient reported outcomes are the best source of data for the main outcome (whether they were feeling better), self-report may have overestimated their actual medication compliance. However, other medication compliance studies reported in the literature also rely on self-report.⁴⁻⁵ We were unable to continue the follow-up after the study period was over. Further evaluations replicating our basic work on the use of automated follow-up and feedback in routine care where all patients are targeted would allow a better assessment of the impact of the follow-up and feedback process. In our secondary study there were also challenges. We had wanted to provide an automated method for analyzing the discordant diagnoses. However that did not prove feasible. Simple lexical or syntactical matching of ICD-9 coding did not provide useful results, for several reasons. First, the discharge diagnosis list was more nearly complete, including chronic conditions important to the patient's overall healthcare trajectory, but not relevant to the index admission. Second, the ICD-9 system has a degree of degeneracy, with multiple codings possible for what clinicians would reasonably consider the same condition. And finally, the ordering of the codes in the datasets was not clearly consistent. Thus, an automated system for identifying discrepancies did not seem viable, and therefore the degree of agreement between the ED admitting diagnoses and final discharge diagnoses was judged manually.

The major issue has been the unexpected difficulty in developing a simple, efficient, and effective method of routinely providing feedback to the original provider in a way that is largely automated and compliant with HIPAA. In addition, a new University restriction on databases containing PHI, restricting them to a single secure server, has decreased ease of access and flexibility in development.

The organization's decision to replace the existing IT infrastructure with a comprehensive EHR created considerable difficulties for the project, in that it diverted resources and will require the redevelopment and revalidation of reporting tools, once complete. Due to the implementation effort, data feeds from the existing system were stopped August 2011. However, in principle, once the implementation is complete the new system should provide a more nearly seamless environment and would better support this effort.

Conclusions

- Between 10% and 20% of ambulatory patients report that their problems are not resolved within a week of their acute care visit and many of the problems persist after three weeks.
- A large proportion of patients do not contact their health care providers when they do not improve as expected.
- Patients and providers appreciate 'closing the feedback loop' and patients who receive follow-up calls are more satisfied with their overall care than those who do not receive the calls.
- Interactive voice response systems are a feasible approach for patient follow-up in ambulatory settings.
- Timely follow-up can be done and it can catch problems at an early stage. Costs for such follow-up can be offset by increased patient care revenue, and early follow-up may avert more costly healthcare expenses, and can potentially improve the quality of care.
- Approximately 10% of Emergency Department residents' diagnoses are discordant with hospital discharge diagnoses. Providing feedback to residents that could address the discordant diagnoses must address the workflow, confidentiality and time constraints inherent in an Emergency Department setting.

References

1. Aspden P, Wolcott J.A., Bootman L, Cronenwett L.R. *Preventing Medication Errors*. Washington, DC: The National Academies Press; 2007.
2. Schiff GD, Bates DW. Can electronic clinical documentation help prevent diagnostic errors? *N Engl J Med*. 2010 Mar 25;362(12):1066-9.
3. Wears RL, Schiff GD. One cheer for feedback. *Ann Emerg Med*. 2005 Jan;45(1):24.
4. Morisky DE, Green W, Levine DM, et al. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care*. 1986;24:67-74.
5. Krousel-Wood M, Islam T, Webber LS, Re RN, Morisky DE, Muntner P. New medication adherence scale versus pharmacy fill rates in seniors with hypertension. *Am J Manag Care*. 2009 Jan;15(1):59-66.
6. Haas JS, Iyer A, Orav EJ, Schiff GD, Bates DW. Participation in an ambulatory e-pharmacovigilance system. *Pharmacoepidemiol Drug Saf*. 2010 Sep;19(9):961-9.
7. Gans N, Koole G, Mandelbaum A. *Telephone Call Centers: Tutorial, Review, and Research Prospects, Manufacturing & Service Operations Management*, Vol. 5, No. 2, pp. 79-141
8. Fleiss, J. L. *Statistical methods for rates and proportions*. New York: John Wiley, 1981.

List of Publications and Products

Peer Reviewed Journal Submissions

Willig, JH, Krawitz, M, Panjamapirom, A, et al. Closing the feedback loop: An interactive voice response system to provide follow-up and feedback in multiple primary care settings. Submitted to *International Journal of Medical Informatics* 07/08/11. Currently under revision and will be submitted to a different journal.

Houser, SH, Ray, MN, Maisiak, R, et al. Patient Satisfaction with interactive voice response follow-up calls. Submitted to *Am J Manag Care* 08/15/11. Currently being revised for resubmission.

Poster and Oral Presentations

Houser SH, Ray MN, Maisiak RM, et al. Do interactive voice response systems as a patient follow up approach increase patient satisfaction? Abstract and poster presentation at AMIA Spring Conference, Public Health Informatics; 2011 May 26-27; Orlando, FL.

Maisiak R, Ray MN, Houser S, et al. The general medication adherence (GMA) scale: Psychometric analysis of a new scale for the primary care setting. Poster presentation at SBM Annual Meeting; 2011 April 29; Washington DC.

Berner ES, Webinar presentation to Creighton. Closing the feedback look in ambulatory care, 2011 March; Omaha, NE.

Berner ES, Burkhardt J, Houser S, et al. Closing the feedback loop to improve diagnostic quality. Presentation at AHRQ HIT Grantees Meeting; 2010 June; Bethesda MD.

Ray MN, Willig J, Cohen M, et al. Follow-up phone calls to improve patient safety in primary care: Issues encountered and lessons learned. Poster presentation at AHRQ HIT Grantees Meeting; 2010 June; Bethesda, MD.

Panjamapirom A, Ray MN, English TM, et al. Patients interest in closing the feedback loop. Poster presentation at AHRQ Annual Conference; September 2009; Bethesda MD.

Berner ES, Ray MN, Schiff GD, et al. Closing the Feedback Loop to Improve Diagnostic Quality. Poster and abstract at AHRQ Annual HIT Annual Conference; 2008 September 7-10; Bethesda MD.

Ray MN, Berner ES, Schiff GD, et al. Closing the Feedback Loop to Improve Diagnostic Quality. Poster presentation at Diagnostic Errors in Medicine Conference; 2008 June; Phoenix AZ.

Inventions

Clinical Documentation and Notification Application (CDNA)—Currently under review by UAB Research Foundation for possible pursuit of patent. Disclosed to HHS on 11/4/2011.

Data Collection Instruments

Patient Outcome Questionnaires (including those for telephone interviews and those for use with IVR)

General Medication Adherence Scale

Patient Satisfaction Survey

Provider Satisfaction Survey