

Human and System Errors, Using Adaptive Turnaround Documents to Capture Data in a Busy Practice

Stephen M. Downs, M.D., M.S., Aaron E. Carroll, M.D., M.S.,
Vibha Anand, M.S., Paul G. Biondich, M.D., M.S.

Children's Health Services Research, Indiana University School of Medicine,
Regenstrief Institute, Inc., Indianapolis, IN

Capturing coded clinical data for clinical decision support can improve care, but cost and disruption of clinic workflow present barriers to implementation. Previous work has shown that tailored, scannable paper forms (adaptive turnaround documents, ATDs) can achieve the benefits of computer-based clinical decision support at low cost and minimal disruption of workflow. ATDs are highly accurate under controlled circumstances, but accuracy in the setting of busy clinics with untrained physician users is untested. We recently developed and implemented such a system and studied rates of errors attributable to physician users and errors in the system. Prompts were used in 63% of encounters. Errors resulting from incorrectly marking forms occurred in 1.8% of prompts. System errors occurred in 7.2% of prompts. Most system errors were failures to capture data and may represent human errors in the scanning process. ATDs are an effective way to collect coded data from physicians. Further automation of the scanning process may reduce system errors.

INTRODUCTION

Providing clinical decision support to, and capturing clinical data directly from physicians at the point of care remains one of the greatest challenges in medical informatics. Electronic medical records and computer based decision support have been shown to improve quality of care,¹⁻³ but systems have not been widely adopted, in part, because of cost and workflow disruption. In addition, decision support is generally built into noting or order writing applications. In many cases, this may be too late because noting and order writing happen after the patient encounter.

In 1995, Downs and colleagues described the use of tailored, scannable paper forms to provide decision support to physicians and capture coded data.⁴ Over several years, the system collected data on tens of thousands of patients.⁵ Shiffman and colleagues also demonstrated that non-tailored scannable forms could be used effectively to capture structured data in a busy pediatric practice.⁶

In 2002 and 2003, our group examined the accuracy of data captured on tailored, scanned paper forms, dubbed adaptive turnaround documents (ATDs), used in a pediatrics clinic.^{7,8} In the latter study, we examined the accuracy of optical character recognition (OCR) and optical mark recognition (OMR) of forms completed by nursing staff (OCR) and families (OMR). We found that of 1309 digits recorded on 221 forms, OCR was 98.6% accurate. OMR of 3131 yes or no answer bubbles on 176 forms was 99.2% accurate. However, we also found that 28.5% of these forms required manual verification before the data could be recorded in the database.

In the above study, digits were entered on the form by a small number of trained clinic staff. Bubble answers were completed by parents, and the manual verification of the scanned images was done by one investigator in a controlled environment. Data capture from ATDs completed by physicians and scanned in a production system pose a new set of challenges. First, physicians are notorious for being sloppy in hand written documentation, presumably because they write quickly. In addition, physicians are often the most difficult to train in system use because of their time constraints. This is exacerbated in an academic setting where learners, such as medical students and residents, come and go, spending relatively brief times in clinic. Finally, verification of scanned forms, when the software cannot interpret a character or a mark, must be done by clinic staff. The accuracy of ATDs used by untrained clinicians in busy clinic settings has not been demonstrated.

We have developed a complete pediatric preventive services system based on ATD technology.⁹ The Child Health Improvement through Computer Automation (CHICA) system has been in operation at the pediatric primary care clinic of Wishard Hospital in Indianapolis since November 5, 2004. In this clinic, ATD forms are completed by physicians who have had no formal training in the scanning technology. The purpose of this study was to measure the frequency of use of the ATD reminders and to see how errors by the physician users and by

the system itself would affect the accuracy of data collection.

METHODS

The CHICA System

CHICA (Child Health Improvement through Computer Automation) is a computer based decision support and electronic record system for pediatric preventive care. CHICA is used as a front end to the Regenstrief Medical Record System.¹⁰ However, it is designed to work as a stand alone application or together with another clinical information system.

CHICA's primary user interface consists of two ATDs⁷ that collect the handwritten responses to dynamically generated questions and clinical reminders while easily integrating into care workflows. To determine what information needs to be printed on each ATD, CHICA employs a library of Arden Syntax¹¹ rules that evaluate the underlying Regenstrief and CHICA databases. Since time constraints limit the number of topics that can be addressed feasibly in a given patient encounter, CHICA also employs a global prioritization scheme¹² which limits the printed content to what's most relevant.

The process begins when registration HL7 messages from our clinic appointment system cue CHICA to begin generating the first of two ATDs. This "pre-screening" form is designed to capture data immediately prior to the provider encounter from both nursing staff and patients. This form has a section for nurses to enter vital signs, and also contains the 20 most important questions to ask a child's family at a particular visit.

Answers extracted from this form are analyzed alongside previously existing data to generate the content for the second ATD. This provider worksheet (PWS) contains reminders which provide varying levels of patient detail based on the information collected before the encounter. (Figure 1) Each of these reminders consists of a "stem" which explains the reason for the prompt and between one and six check boxes with which the physician can document his or her response to the prompt. To the right of each check box is a small "erase box." If this box is filled in, the system treats the corresponding box as not checked. The physicians received no formal training in the use of the CHICA forms.

Figure 1. A physician worksheet generated by the CHICA system. The eight prompts in the center of the page include check boxes that are scanned and interpreted by the TeleForms software.

The PWS forms are typically scanned into the computer by nursing or clerical staff after the clinic session. These staff members are responsible for verifying the system's interpretation of the forms when the computer interpretation is ambiguous.

Figure 2. An example of a prompt generated by the CHICA system and correctly completed by the physician. As marked, this prompt should write one observation: Any Family Members Smoke = Yes Not Ready to Quit.

All data collected from both of these forms are ultimately stored as coded clinical observations which are used to drive decision support and can be analyzed retrospectively. (Figure 2) CHICA also stores a Tagged Image File Format (TIFF) image of the worksheets. In the current environment, the paper is the note of record. However, eventually, the TIFF and the stored observations may be digitally signed and stored electronically as the note of record.

RESULTS

CHICA is used by 12 attending physicians, 16 resident physicians and a variety of medical students on their outpatient pediatrics rotations.

Evaluation of System Accuracy

From the CHICA database, we pulled a random sample of 200 TIFF images of PWS forms scanned between December 1, 2004 and February 28, 2005. We also extracted all of the observations recorded from those scans. Two investigators, working independently, reviewed the forms and the observations recorded from them, looking for (1) errors of omission, in which the physician intended to check a box, but the corresponding observation(s) were not recorded in the medical record; and (2) errors of commission in which the physician did not intend to check a box, but the corresponding observation(s) were recorded in the database. These errors were further divided into those in which the physician marked the form incorrectly and those in which the physician marked the form correctly, but the system somehow wrote or failed to write an observation in error. Thus, we counted four types of errors: (1) errors of commission caused by the physician, (2) errors of omission caused by the physician, (3) errors of commission caused by the system, and (3) errors of omission caused by the system.

After the two investigators had identified all of the errors in 200 forms, their results were compared, and disagreements were resolved by consensus. We calculated a kappa statistic for agreement above chance.

Because each physician prompt has up to six check boxes, and different combinations of checked boxes can write any number of observations, choosing a denominator for the error frequency calculation was a challenge. We chose to use, as the denominator, the number of prompts used by the physician, i.e., prompts in which the physician checked at least one box. This metric may inflate the error rate relative to other measures of scanning accuracy because scanning each prompt may require optical mark sensing of up to 12 boxes (six answer boxes and six "erase" boxes). However, given the variation in the number of responses per prompt, observations per check box, and prompts completed per form, this seemed the most consistent and meaningful denominator.

Agreement Between Evaluators

There was a high level of inter-rater reliability between evaluators. The kappa statistic was 0.82, corresponding to excellent agreement above chance.

Use of the Decision Support Reminders

Each of the PWS forms can present up to eight reminders. Most forms have eight, but in a few circumstances, particularly among children seen recently or frequently and with few risk factors, fewer than eight prompts may appear on the form. We found that at least one prompt was used by the physician on 63% of the forms. On forms where at least one prompt was used, the average number of prompts marked by the physician was 4.3, or somewhat over half. Certain physicians were more likely than others to use the forms. Likewise, when the forms were used, some prompts were completed more frequently than others.

It was our anecdotal observation that younger physicians were more likely to use the prompts than more experienced clinicians. We also observed that the prompts that addressed fairly pedestrian topics like sleep patterns or feeding schedules were more likely to be used than more sensitive or work intensive issues like maternal depression or screening tests. Some clinicians reported that they did not always agree with the recommendations in the prompts despite the fact that they are based on standard, authoritative guidelines. None the less, rates of use of the prompts appears to be increasing with time.

Accuracy of ATD Completed by Physicians

Table 1 shows the rates of each of the four types of errors. Among the 542 prompts used by physicians, there were 49 errors for a 9% error rate. The vast majority of these (7.2%) were errors of omission caused by the system. In other words, the physician had checked the forms correctly, as in Figure 2, but the system had not recorded the corresponding observations. We observed that many of these system errors of omission occurred in groups on the same forms, suggesting that there was an error in the scanning of the entire form. Furthermore, nearly half of these errors occurred on forms scanned on two particular days. When errors of omission were due to physician error, they were the result of checking outside the box (Figure 3).

Table 1. Error rates for errors of omission (in which an observation was mistakenly not recorded in the database) caused by physician or by system, and errors of commission (in which an observation was written that was not intended) caused by physician or by system.

	Errors of Omission	Errors of Commission	Totals
Error by Physician	0.7%	1.1%	1.8%
Error by System	7.0%	0.2%	7.2%
Totals	7.7%	1.3%	9.0%

Errors of commission were very uncommon, and these were almost never caused by the system. They tended to occur when the physician made stray marks on the form or made large marks that went over more than one box (Figure 4).

Figure 3. Error of omission caused by the physician checking the form outside the designated box. The check mark, apparently intended to indicate “city water,” falls outside of any of the check boxes.

Fluoride prevents tooth decay. City water is fluoridated, well water is usually not. Fluoride dosage should be 1/2 of recommended if child is in daycare/school with well water at home.

<input checked="" type="checkbox"/> All H2O from well or bottled->	<input type="checkbox"/> 0.50 mg Fluoride qD
<input type="checkbox"/> city water	<input type="checkbox"/>
<input type="checkbox"/> Comb city H2O & well/bottled->	<input type="checkbox"/> 0.25 mg Fluoride qD

Figure 4. Error of commission caused by physician incorrectly marking the form. The mark, apparently intended to indicate “city water,” inadvertently marked two boxes.

Fluoride prevents tooth decay. City water is fluoridated, well water is usually not. Fluoride dosage should be 1/2 of recommended if child is in daycare/school with well water at home.

<input type="checkbox"/> All H2O from well or bottled->	<input type="checkbox"/> 1 mg Fluoride qD
<input checked="" type="checkbox"/> city water	<input type="checkbox"/>
<input checked="" type="checkbox"/> Comb city H2O & well/bottled->	<input type="checkbox"/> 0.50 mg Fluoride qD

DISCUSSION

In the first four months of use of the CHICA system, we have seen adoption of the physician forms in 63% of encounters. When the forms are used, just over half of the physician prompts are utilized. This is encouraging progress. However, the full benefit of the system will not be realized until the system is universally adopted. Our experience with CHICA’s predecessor⁴ has shown that adoption of an ATD system is dependent on a multitude of system and clinic factors.¹³ Anecdotally, we observed that physicians’ adoption of ATDs improves when the physicians begin to see that the data they provided at earlier visits are utilized in subsequent prompts. We

also plan to encourage use of the system by feeding back aggregate data obtained by CHICA.

CHICA showed moderately good accuracy in a “real world” clinic. Errors caused by physicians, who were never formally trained to use the system, were surprisingly uncommon. In fact, physicians’ overall accuracy was 98.2%. Looking separately at the types of errors, the positive and negative predictive value of data entered by the physicians was 99.3% and 98.9% respectively. This is significantly better than more traditional sources of physician generated data such as ICD-9 billing codes. The positive predictive value of this information is between 86% and 94%.^{14, 15} With additional training, these rates may become negligible. This speaks to one of the biggest advantages of the ATD concept: that the paper interface and the check-box format are so familiar to clinicians that they can intuitively complete the forms in a way that captures meaningful clinical data. Although the technology could be moved to a mobile (palmtop or tablet) computing environment, paper may still be preferred by clinicians. This would be interesting future work.

It is also reassuring that the system is extremely unlikely to cause errors of commission. However, the 7.2% frequency of errors of omission, when the physicians had marked the forms correctly, deserves particular attention. There are three factors to consider in interpreting this number. First, we chose to evaluate error rates in terms of errors per prompt used by the physician. But, in fact, each prompt represents up to 16 OMR fields or check boxes that must be recognized. Thus, the error rate is inflated relative to the usual way OMR accuracy is reported.

None the less, the 7.2% error rate seems too high. One possible reason is that CHICA’s physician forms represent the most complex application to date of the Verity software suite (private communications with Verity). The software was never intended to support the complexity of the PWS. Not surprisingly, creating a process for scanning a large volume of these forms and confirming their accuracy in the setting of a very busy practice, with no additional personnel presents a challenge.

Because system errors tended to cluster on individual forms, it is likely that the entire form was misread. In practice, we have seen that several forms require verification at the time of scanning, similar to the need for verification seen in our earlier work.⁷ In the production system, this is done under time pressure by clinic staff. It is likely that this is the point at which errors are made. The fact that many of the

errors happened on a couple of specific days supports this hypothesis.

Ironically, we have found that when verification is required, the software has usually made a correct guess at the OMR value. Thus, the verification by harried staff may actually be introducing errors. We are planning experiments to bypass the verification step and see if, by eliminating the human element, the data quality actually improves.

Despite the challenges of perfecting workflow issues, the ATD model achieves capture of coded clinical data at low cost, with minimal disruption of workflow and with essentially no user training. Because the accuracy of data capture is comparable to other common electronic health data (e.g., electronic claims) we anticipate this will be a practical form of data capture and clinical support for busy practices.

References

1. Dexter, P.R., et al., *A computerized reminder system to increase the use of preventive care for hospitalized patients.[see comment]*. New England Journal of Medicine, 2001. **345**(13): p. 965-70.
2. Hunt, D.L., et al., *Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review*. JAMA, 1998. **280**(15): p. 1339-46.
3. Harris, R.P., et al., *Prompting physicians for preventive procedures: a five-year study of manual and computer reminders*. American Journal of Preventive Medicine, 1990. **6**(3): p. 145-52.
4. Downs, S., J. Arbanas, and L. Cohen, *Computer supported preventive services for children*. Proceedings of the Nineteenth Annual Symposium on Computer Applications in Medical Care, 1995: p. 962.
5. Downs, S.M. and M.Y. Wallace, *Mining association rules from a pediatric primary care decision support system*. Proceedings / AMIA, 2000. **Annual Symposium.**: p. 200-4.
6. Shiffman, R.N., C.A. Brandt, and B.G. Freeman, *Transition to a computer-based record using scannable, structured encounter forms*. Archives of Pediatrics & Adolescent Medicine, 1997. **151**(12): p. 1247-53.
7. Biondich, P.G., et al., *Using adaptive turnaround documents to electronically acquire structured data in clinical settings*. AMIA, 2003. **Annual Symposium Proceedings/AMIA Symposium.**: p. 86-90.
8. Biondich, P.G., et al., *A modern optical character recognition system in a real world clinical setting: some accuracy and feasibility observations*. Proceedings / AMIA, 2002. **Annual Symposium.**: p. 56-60.
9. Anand, V., et al. *Child Health Improvement through Computer Automation: The CHICA System*. in *MEDINFO 2004*. 2004. San Francisco: IOS Press, IMIA, Amsterdam.
10. McDonald, C.J., et al., *The Regenstrief Medical Record System: 20 years of experience in hospitals, clinics, and neighborhood health centers*. MD Computing, 1992. **9**(4): p. 206-17.
11. Jenders, R.A., et al., *Medical decision support: experience with implementing the Arden Syntax at the Columbia-Presbyterian Medical Center*. Proceedings - the Annual Symposium on Computer Applications in Medical Care, 1995: p. 169-73.
12. Downs, S.M. and H. Uner, *Expected value prioritization of prompts and reminders*. Proceedings / AMIA, 2002. **Annual Symposium.**: p. 215-9.
13. Travers, D.A. and S.M. Downs, *Comparing user acceptance of a computer system in two pediatric offices: a qualitative study*. Proceedings / AMIA, 2000. **Annual Symposium.**: p. 853-7.
14. Kiyota, Y., et al., *Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records*. American Heart Journal, 2004. **148**(1): p. 99-104.
15. Losina, E., et al., *Accuracy of Medicare claims data for rheumatologic diagnoses in total hip replacement recipients*. Journal of Clinical Epidemiology, 2003. **56**(6): p. 515-9.

Acknowledgements

This work was supported, in part, by a grant from the Riley Children's Foundation.

Address Correspondence to:

Children's Health Services Research
Indiana University School of Medicine
699 West Drive, RR330
Indianapolis, IN 46077