Towards Evaluation of a Medical Case-Based Decision Support System

Cindy Marling[1], Stan Vernier[1], Tessa Cooper[1],
Jay Shubrook[2] and Frank Schwartz[2]

[1] School of Electrical Engineering and Computer Science
Russ College of Engineering and Technology
Ohio University, Athens, Ohio 45701, USA
marling@ohio.edu, sv193702@ohio.edu, tc273702@ohio.edu

[2] Appalachian Rural Health Institute, Diabetes and Endocrine Center
College of Osteopathic Medicine
Ohio University, Athens, Ohio 45701, USA
shubrook@ohio.edu, schwartf@ohio.edu

Abstract. A clinical research study is underway to evaluate a medical case-based decision support system in the domain of diabetes management. Thirty patients with type 1 diabetes on insulin pump therapy are participating in the study to evaluate the 4 Diabetes Support System. System evaluation is especially important in medical domains, because systems must demonstrate positive impact on patient outcomes if they are to be used in practice. This study follows a preliminary system evaluation and precedes a full randomized clinical trial to quantify clinical outcomes. An overview of the 4 Diabetes Support System, the evaluation study protocol, and preliminary results of the evaluation are presented.

1 Introduction

In a medical domain, the ultimate success of a case-based decision support system is determined by its impact on patient outcomes. Patients who use a system should have some measurable advantage over those who do not, in terms of health, longevity, quality of life and/or cost of health care. As explained in [1], measuring the clinical outcomes of a system requires a randomized clinical trial, in which some patients use the system and others do not. However, because such a trial is expensive, in terms of time and human resources, the evaluation process may be phased. In the first phase, it is important to thoroughly evaluate the system in terms of its accuracy and usability. This is essential not only to maximize the success of the later trial, but also to minimize any potential negative impact on patients participating in the trial.

The Data-Driven Diabetes Decision Support (4 Diabetes Support) System aims to help physicians manage patients with type 1 diabetes on insulin pump therapy. These patients need to vigilantly monitor their blood glucose levels,

keeping them as close to normal as possible, to avoid serious diabetic complications. Helping these patients make therapy adjustments to combat problems in blood glucose control is a data intensive, time-consuming task for physicians. A preliminary clinical study was conducted in which a prototype of the 4 Diabetes Support System was constructed [2–4]. Because preliminary results were encouraging, a second clinical study is underway to evaluate and enhance this prototype, prior to the conduct of a randomized clinical trial. This paper briefly reviews the 4 Diabetes Support System, describes the evaluation protocol, and presents preliminary findings of the evaluation.

2 The 4 Diabetes Support System

This section gives a brief overview of the 4 Diabetes Support System. The 4 Diabetes Support System prototype was built during a preliminary clinical study in which 20 patients with type 1 diabetes on insulin pump therapy participated. Fifty cases of problems in blood glucose control, with their associated solutions and clinical outcomes, were compiled into a central case base. Specific problems are rich in context and vary widely. However, they revolve around hyperglycemia, hypoglycemia, and fluctuations between the two. Hyperglycemia, or high blood glucose, is responsible for serious long-term complications of diabetes, including blindness, neuropathy, and heart failure. Hypoglycemia, or low blood glucose, may result from insulin treatment to control hyperglycemia. Its effects are more immediate, including weakness, confusion, dizziness, sweating, shaking, and, if not treated promptly, loss of consciousness or seizure. Physicians propose adjustments to therapy as solutions to combat these problems. Adjustments are changes involving insulin, food and/or exercise.

An overview of the system operation can be seen in Figure 1. Patients enter daily glucose, insulin and life event data into an Oracle database via a Web browser. Data from a continuous glucose monitoring system (CGMS) is uploaded directly from the patient's monitoring device to the database. The situation assessment module analyzes the data to detect the problems in glucose control that a patient is experiencing. This module contains 12 problem detection routines, as listed in Figure 2. After detection, these problems are presented to the physician for review. The physician selects the most critical problem or problems for the patient. The next step performed by the system is to retrieve the closest matching case for each selected problem. A standard two-step retrieval method, with a nearest neighbor algorithm, is used, as reported in [3].

Results of the preliminary evaluation were encouraging, but problems were also identified. Participating patients completed an exit survey, which indicated acceptance of the concept of automated decision support. The time required for data entry was noted as a potential impediment to use, however. It took patients between 15 and 60 minutes per day to enter data, and some patients who did not complete the entire protocol cited the time required for data entry as a reason. A panel of three physicians and one advance practice nurse specializing in diabetes reviewed a random sample of problems detected by the situation assessment

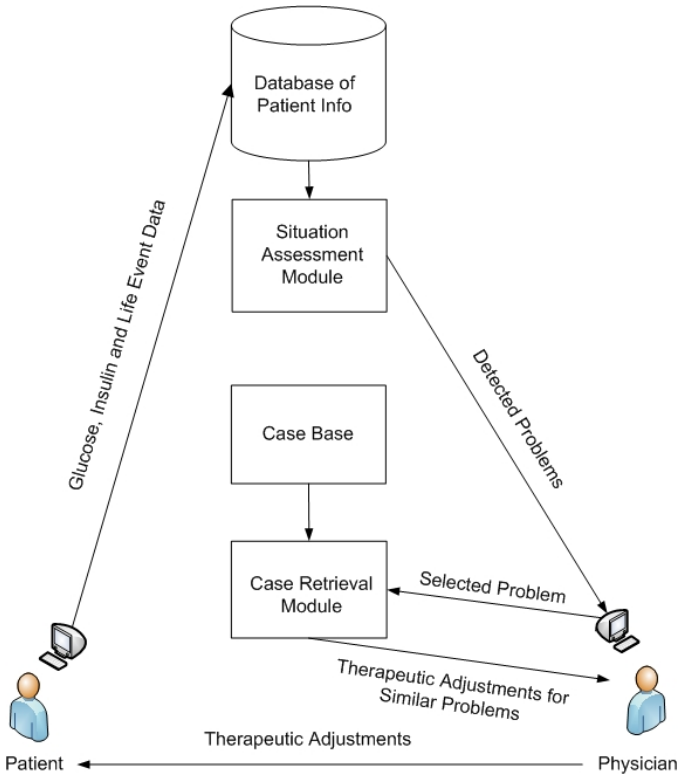

Fig. 1. Overview of System Operation

1. Over-correction for hypoglycemia
2. Post-exercise hypoglycemia
3. Possible pump or infusion set malfunction
4. Over-correction for hyperglycemia
5. Pre-waking hypoglycemia
6. Over-bolus for a meal
7. Hyperglycemia upon awakening
8. Hypoglycemia upon awakening
9. Pre-meal hyperglycemia
10. Pre-meal hypoglycemia
11. Post-meal hyperglycemia
12. Post-meal hypoglycemia

Fig. 2. Situation Assessment Routines

module. They considered the problems to be correctly identified 77.5% of the time, and thought it would be useful to call the problems to the attention of a physician 90% of the time. Leave one out testing was performed to evaluate the case retrieval module. Three diabetes specialists reviewed a random sample of cases with their nearest neighbors. They considered the cases to be similar 80% of the time, and thought that the solutions stored in matching cases would be helpful in solving the original problems 70% of the time. It was noted that not all cases had usefully matching nearest neighbors, and the need to expand the case base was identified. A concern was raised that the test data used was for patients who had contributed data for building the system.

3 The Evaluation Study Protocol

The evaluation study aims to more fully evaluate the performance of the 4 Diabetes Support System prototype, identify future development needs, and reduce the data entry burden for patients. Thirty patients with type 1 diabetes on insulin pump therapy were recruited into the evaluation study. Patients who participated in the initial study were excluded, so that the system could be tested on data for patients whose problems had not been included in the case base.

Each patient visited the Appalachian Rural Health Institute Diabetes and Endocrine Center for an initial visit. At this visit, following informed consent, each patient completed a brief multiple choice inventory designed to gauge the patient's perception of his or her current diabetes control and its impact on quality of life. The HbA1c, a measure of long-term blood glucose control, was obtained. Next, the patient provided background information about his or her current health status, diabetes treatment, and typical daily routines. The times the patient normally awakes, goes to sleep, works or attends school, exercises and eats were recorded.

Each patient was then shown how to automatically transfer data from his or her insulin pump to Medtronic's CareLink data management Web site [5]. CareLink is a commercially available data collection and visualization tool that is free to patients who use Medtronic insulin pumps. It consists of separate, but connected, modules for patients and physicians, CareLink Personal [6] and CareLink Pro [7], respectively. Data is wirelessly transmitted from the patient's pump to CareLink Personal via a USB download device. This data can then be viewed graphically and in log form by the patient in CareLink Personal and by the patient's physician in CareLink Pro. The patient was asked to send data to CareLink once a week for the next four weeks. This data was later extracted, de-identified, and incorporated into the experimental database.

To ease the data entry burden, patients were not asked to input their life events on a daily basis. The simplifying assumption was made that the patient slept, worked, exercised and ate at the times given by their typical daily schedule. Patients were asked to email their physicians with any unusual life events they felt might be impacting their blood glucose levels.

Once a week, data was aggregated and presented for physician review. Problems identified by the situation assessment module were discussed at each weekly review meeting. For each problem detected by the system, the participating patient's physician was asked if it was (a) correct and (b) useful to detect that problem for the patient. Problems that were recognized by the physicians but were not automatically detected by the system were also recorded. These represent future development needs.

Later, the system's ability to retrieve useful cases for the detected problems was tested. This was not done at weekly meetings because solutions are normally sought for only a patient's most critical problems. For thoroughness of evaluation, however, solutions were sought for each type of problem detected for each patient. These were initially reviewed by the knowledge engineering team and are awaiting full review by the physicians.

4 Preliminary Results of the Evaluation

Situation assessment data has been tabulated for the first 20 patients who participated in the evaluation study. Eighteen of these patients completed the full five weeks of observation. Of the two who did not, one was mistakenly enrolled with an incompatible pump type and the other switched pumps midway through the study, losing some of the data. No patients withdrew due to the time required to provide the requested data. This was viewed as a positive finding, because in the preliminary study, 40% of the patients did not complete the entire protocol. Some cited the time required to enter daily data as an impediment to their participation.

The 12 situation assessment routines, listed in Figure 2, found a total of 222 possible problems for the 18 patients. Of the 222 problems detected, 189 were evaluated by the physician of the patient for whom the problem was found. The physicians concluded that 186 of the 189, or 98.4% of the evaluated problem detections were correct and three were incorrect. In regards to whether the detections were useful, physicians rated 181, or 95.8% as useful, six as not useful and two as possibly useful. The details compiled by problem detection routine are seen in Table 1, and by patient in Table 2.

The correctness of the problem detections appears to be better than in the preliminary study. This may be due, in part, to bug fixes, and in part to differences in the way the correctness was evaluated. Only the patient's own physician, who was most familiar with the patient's actual problems, was asked to verify correctness of the problems detected.

It should also be noted that there is no way to accurately determine the number of problems experienced by patients that were not detected by the system. Physicians do not have time to manually detect the large number of problems that could occur in patient data. That is one of the driving factors behind providing automated problem detection in the first place. However, there is reason to believe that many actual problems were missed. The problem detection routines found fewer problems per patient than were found in the preliminary study.

		Correct to Detect			Useful to Detect			
Routine	Total	Yes	No	Maybe	Yes	No	Maybe	Not Evaluated
1	0	0	0	0	0	0	0	0
2	18	15	1	0	15	1	0	2
3	21	10	2	0	10	2	0	9
4	7	7	0	0	7	0	0	0
5	1	0	0	0	0	0	0	1
6	1	1	0	0	0	0	1	0
7	18	18	0	0	18	0	0	0
8	48	40	0	0	40	0	0	8
9	22	18	0	0	15	3	0	4
10	51	46	0	0	45	0	1	5
11	2	2	0	0	2	0	0	0
12	33	28	0	0	28	0	0	5
Totals	222	186	3	0	181	6	2	33

Table 1. Results by Situation Assessment Routine

		Correct to Detect			Useful to Detect			
Patient	Total	Yes	No	Maybe	Yes	No	Maybe	Not Evaluated
1	9	9	0	0	9	0	0	0
3	2	1	0	0	1	0	0	1
4	9	9	0	0	9	0	0	0
5	6	6	0	0	6	0	0	0
7	12	10	0	0	9	0	1	2
8	9	9	0	0	9	0	0	0
9	19	19	0	0	19	0	0	0
10	9	7	0	0	6	0	1	2
11	21	12	0	0	12	0	0	9
12	45	41	3	0	41	3	0	1
13	9	4	0	0	4	0	0	5
14	0	0	0	0	0	0	0	0
15	14	13	0	0	13	0	0	1
16	15	11	0	0	11	0	0	4
17	16	12	0	0	9	3	0	4
18	9	8	0	0	8	0	0	1
19	3	1	0	0	1	0	0	2
20	15	14	0	0	14	0	0	1
Totals	222	186	3	0	181	6	2	33

Table 2. Results by Patient

An average of 12.3 problems per patient were detected in this study versus 29.3 problems per patient in the earlier study. Even pro-rating for the difference in the length of the studies, this study found 50.5% fewer problems per patient [8].

Lack of life-event data appears to account for many missed problem detections. In the preliminary study, patients provided additional life-event data, including their actual daily work and exercise schedules. Because it was time-consuming for patients to provide this data, the current study was designed to determine the impact of using typical daily schedules instead. For all four of the meal related routines and the low after exercise routine, it is disadvantageous not having a patient's actual schedule. Assumptions are made that the patient is exercising every time they plan to, and that recorded carbohydrate intakes within one hour of their usual meal time are actual meals. Carbohydrate intakes recorded at other times are assumed to be snacks. This prevents the system from detecting pre-meal, post-meal, and post-exercise problems when a patient's schedule varies from normal.

While physicians have yet to fully evaluate the case retrieval module of the system, initial observations indicate that the system is not finding matching cases as well as it did in the preliminary study. While CBR, in general, may be robust to missing values, the missing life event data appears to have provided essential context for describing, differentiating, and comparing cases. While generally matching cases are found, more specifically matching cases may be overlooked. Another problem is that the 12 problem detection routines do not account for all of the problems in the case base. Rather, they were developed to account for the most common problems. This effectively makes the case base smaller, as some cases are never good matches for any of the problems detected. While these are negative findings, they do clarify the needs to acquire more life event data, expand the case base, and develop additional problem detection routines.

The system demonstrated its potential benefit when a participating patient was hospitalized with diabetic ketoacidosis. This patient's pump had malfunctioned, so that his insulin was not delivered, and his blood glucose rose. The patient knew that his blood glucose was high, but he did not know that the pump was not delivering the insulin with which he tried to correct his hyperglycemia. He went into diabetic ketoacidosis and experienced an acute coronary event with a silent heart attack. Running retroactively, the system was able to detect the pump problem eight hours before the patient was admitted to the hospital. Had the system been running in real time, it might have been possible to alert the patient to this problem in time for him to correct it.

5 Future Work

The immediate task at hand is to finish the analysis for the current evaluation. Situation assessment data for the remaining ten patients still needs to be tabulated and analyzed. The similarity and usefulness of the cases retrieved needs to be evaluated by a panel of physicians.

Another clinical study is planned to significantly grow the case base and to address the other issues identified during the evaluation. Twenty-eight patients with type 1 diabetes on insulin pump therapy will take part. These patients will automatically upload data from their insulin pumps using Medtronic's CareLink software. However, they will supplement this data with life-event data not stored in their pumps to provide a fuller context for problem solving. They will do this using a shortened and simplified version of the Web browser based interface used in the preliminary study. The longer range goal is for patients to use the technology in their own medical devices and/or cell phones to facilitate data entry.

Once the accuracy and usability of the system is further validated, the next step will be to conduct a multi-site randomized clinical trial. Pre- and post-values of HbA1c, a measure of long-term glucose control, will be used to gauge the efficacy of system use. If measurable improvement in patient outcomes can be demonstrated, the 4 Diabetes Support System may advance beyond the research laboratory into clinical practice.

6 Related Research

This work is anchored within the framework of CBR in the Health Sciences, which is in turn part of a long tradition of research in AI in Medicine [9]. As noted in [1], such work is driven by both the desire to advance the scientific knowledge of AI and CBR and by the real-world needs of patients and health care professionals. For the past six years, workshops on CBR in the Health Sciences have been held at every International and European Conference on Case-Based Reasoning (ICCBR and ECCBR). Several good overview papers have been written on work in the field, including [10–12].

Diabetes management was first identified as a fruitful domain for CBR research by the Telematic Management of Insulin-Dependent Diabetes Mellitus (T-IDDM) project [13–15]. This was a telemedicine project that aimed to remotely monitor and support patients in maintaining good glucose control. T-IDDM was a hybrid system that relied primarily on rule-based reasoning and a probabilistic model of the effects of insulin on blood glucose over time. CBR was integrated to tune rule parameters to optimize advice for patients.

Diabetes shares much in common with other chronic diseases that can not be cured but must be managed or treated over time. Chronic disease management involves consideration of time-varying data, patient variability, and individual patient preferences and needs. Related work has been conducted in the domains of psychiatric eating disorders [16], stem cell transplantation follow-up care [17], end-stage renal disease [18], and Alzheimer's disease [19].

7 Summary and Conclusions

A clinical study of 30 patients with type 1 diabetes on insulin pump therapy was designed to evaluate the 4 Diabetes Support System prototype. This is an

intermediate step between the preliminary evaluation conducted when the prototype was originally built and the randomized clinical trial needed to measure its impact on patient outcomes. Test data was collected from patients who were not involved in the original study in which the case base was built. This helped to ensure that the system was not overfit to the individual test cases, a potential criticism of leave one out testing. User interface issues identified during the original study were addressed. While results are still being tabulated and analyzed, preliminary results have already identified system strengths, weaknesses, and development needs. Evaluation studies are especially crucial for systems developed in medical domains, because positive impact must be demonstrated before systems can move beyond the research laboratory into clinical use. Conducting evaluations that demonstrate impact is difficult for many reasons, including financial constraints, time constraints, and the rapidly evolving nature of software systems [20]. Additional work is needed to define and document practical system evaluation methodologies suitable for CBR in the Health Sciences.

8 Acknowledgments

The authors gratefully acknowledge support from Medtronic, and from Ohio University through the Russ College Biomedical Engineering Fund, the Osteopathic College of Medicine Research and Scholarly Affairs Committee, and the Appalachian Rural Health Institute Diabetes Research Initiative. We would also like to thank the participating patients, research nurses, and graduate research assistants for their contributions to this work.

References

1. Bichindaritz, I.: Case-based reasoning in the health sciences: Why it matters for the health sciences and for CBR. In Althof, K.D., Bergmann, R., Minor, M., Hanft, A., eds.: Advances in Case-Based Reasoning: 9th European Conference, ECCBR, Berlin, Springer (2008) 1–17
2. Marling, C., Shubrook, J., Schwartz, F.: Towards case-based reasoning for diabetes management. In Wilson, D.C., Khemani, D., eds.: Workshop Proceedings of the Seventh International Conference on Case-Based Reasoning (ICCBR-07), Belfast, Northern Ireland (2007) 305–314
3. Marling, C., Shubrook, J., Schwartz, F.: Case-based decision support for patients with type 1 diabetes on insulin pump therapy. In Althof, K.D., Bergmann, R., Minor, M., Hanft, A., eds.: Advances in Case-Based Reasoning: 9th European Conference, ECCBR, Berlin, Springer (2008) 325–339
4. Schwartz, F.L., Shubrook, J.H., Marling, C.R.: Use of case-based reasoning to enhance intensive management of patients on insulin pump therapy. Journal of Diabetes Science and Technology **2**(4) (2008) 603–611
5. Medtronic MiniMed: CareLink personal therapy management software for diabetes (2009) https://carelink.minimed.com/patient/, accessed April, 2009.
6. Medtronic MiniMed: CareLink Personal Therapy Management Software for Diabetes: User Guide. Medtronic MiniMed, Northridge, California (2008) Available at http://www.minimed.com/pdf/carelink_user_guide.pdf.

7. Medtronic MiniMed: CareLink Pro Therapy Management Software for Diabetes: User Guide. Medtronic MiniMed, Northridge, California (2007) Available at http://www.minimed.com/pdf/UserGuide7335_ENGLISH.pdf.
8. Miller, W.: Problem Detection for Situation Assessment in Case-Based Reasoning for Diabetes Management. Master's thesis, Ohio University, Athens, Ohio (2009)
9. Altman, R.B.: AI in medicine: The spectrum of challenges from managed care to molecular medicine. AI Magazine **20**(3) (1999) 67–77
10. Nilsson, M., Sollenborn, M.: Advancements and trends in medical case-based reasoning: An overview of systems and system development. In: Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference – Special Track on Case-Based Reasoning, Menlo Park, California, AAAI Press (2004) 178–183
11. Holt, A., Bichindaritz, I., Schmidt, R., Perner, P.: Medical applications in case-based reasoning. The Knowledge Engineering Review **20** (2005) 289–292
12. Bichindaritz, I., Marling, C.: Case-based reasoning in the health sciences: What's next? Artificial Intelligence in Medicine **36** (2006) 127–135
13. Bellazzi, R., Larizza, C., Montani, S., Riva, A., d'Annunzio, M.S.G., Lorini, R., Gómez, E.J., Hernando, E., Brugués, E., Cermeno, J., Corcoy, R., de Leiva, A., Cobelli, C., Nucci, G., Prato, S.D., Maran, A., Kilkki, E., Tuominen, J.: A telemedicine support for diabetes management: the T-IDDM project. Computer Methods and Programs in Biomedicine **69** (2002) 147–161
14. Montani, S., Bellazzi, R.: Supporting decisions in medical applications: The knowledge management perspective. International Journal of Medical Informatics **68** (2002) 79–90
15. Montani, S., Magni, P., Bellazzi, R., Larizza, C., Roudsari, A.V., Carson, E.R.: Integrating model-based decision support in a multi-modal reasoning system for managing type 1 diabetic patients. Artificial Intelligence in Medicine **29** (2003) 131–151
16. Bichindaritz, I.: MNAOMIA: Improving case-based reasoning for an application in psychiatry. In: Artificial Intelligence in Medicine: Applications of Current Technologies, Stanford, CA, Working Notes of the AAAI-96 Spring Symposium (1996)
17. Bichindaritz, I., Kansu, E., Sullivan, K.M.: Case-based reasoning in CAREPARTNER: Gathering evidence for evidence-based medical practice. In Smyth, B., Cunningham, P., eds.: Advances in Case-Based Reasoning: 4th European Workshop, Proceedings EWCBR-98, Berlin, Springer-Verlag (1998) 334–345
18. Montani, S., Portinale, L., Leonardi, G., Bellazzi, R.: Applying case-based retrieval to hemodialysis treatment. In McGinty, L., ed.: Workshop Proceedings of the Fifth International Conference on Case-Based Reasoning, Trondheim, Norway (2003)
19. Marling, C., Whitehouse, P.: Case-based reasoning in the care of Alzheimer's disease patients. In Aha, D.W., Watson, I., eds.: Proceedings of the Fourth International Conference on Case-Based Reasoning, ICCBR-01, Berlin, Springer (2001) 702–715
20. Moehr, J.R.: Evaluation: Salvation or nemesis of medical informatics? Computers in Biology and Medicine **32** (2002) 113–125