Prof. Edward Aboufadel
Department of Mathematics
Sabbatical Report for Fall 2011 Sabbatical
January 2012

I wish to thank Grand Valley for providing me the time to pursue a research project in the area of diabetes and mathematics. I feel like I accomplished a lot, and I also had time to reflect on my career so far, and my next steps. This is my report of what I did professionally during my sabbatical, from August to December 2011.

The research project

In my sabbatical proposal, I outlined my plans to use publicly-available continuous glucose monitor (CGM) data to consider two research questions pertaining to type-1 diabetes patients:

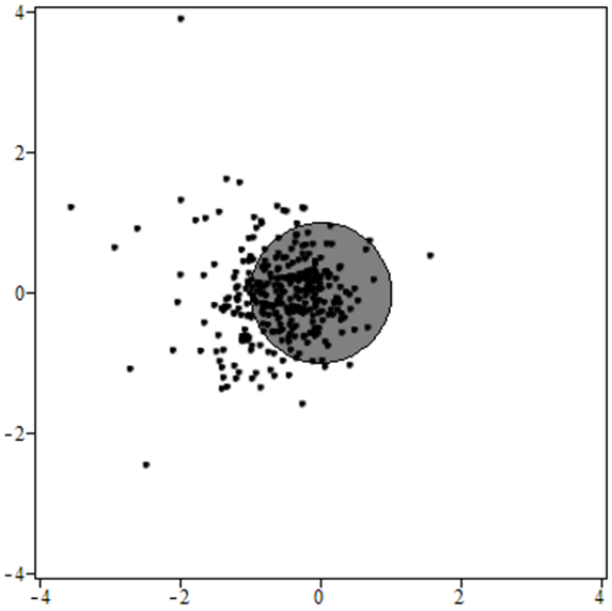
1. Can the dynamics of an individual diabetic patient's blood glucose numbers be used to predict future short-term blood glucose behavior? (For example, is there a set of patients for which two days of relatively stable data implies with high probability a highly variable third day? If so, what counseling can be provided to these patients to adjust their blood glucose management going into the third day?)

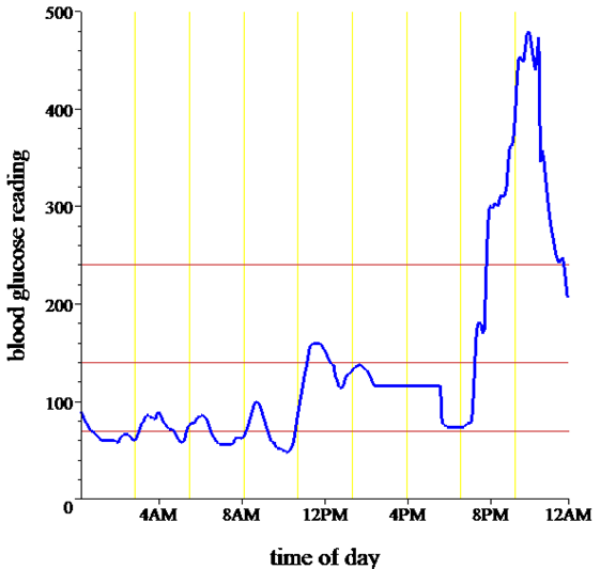
2. Recently, Netflix (the movie rental company) ran a public contest to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences. Successful approaches involved intricate mathematical analyses of the large data set that Netflix made public, using mathematical tools such as singular value decomposition of large matrices. A key idea was to develop measures of similarities between Netflix customers. How can these approaches be adapted to blood glucose data in order to create profiles of glucose behavior which can be matched up with diabetic patients, beyond what was done previously.

During my sabbatical, I pursued these questions, but ended up in a different place than I expected. I started by creating a "CGM comparison engine" to find similarities between the 42,799 individual days of CGM data that I had available. I struggled to find ways to modify the measures of similarity from the Netflix prize solutions, and instead developed a robust method based on the mathematics of a smartphone app called *Shazam*. This app will identify a song after five seconds of the song is collected through the microphone. A key part of the CGM engine is a wavelet filter, which I have used successfully in my scholarship for a decade. Completing the CGM engine by mid-October gave me the tool that I needed to answer some questions.

As far as question #1, the answer turned out to be "not at all". There appears to be no relationship between day X of a diabetic patient's glucose behavior and day X+1. As to question #2, my work led to the development of diagrams that patients and physicians can use to compare their glucose management to other patients. The basic idea is to use wavelet filters to create a

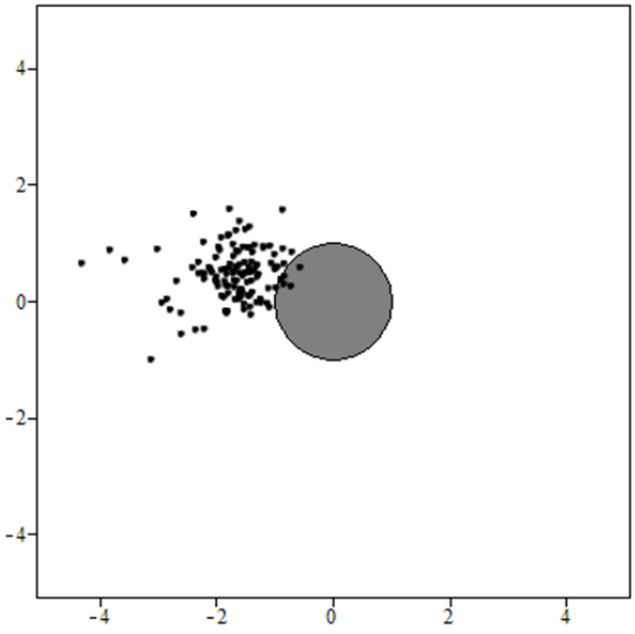
"wavelet signature" for each CGM day, and then to use the signatures to compute a "distance" between CGM days. Treating each CGM day as a point, we can use these distances and a mathematical procedure called "multi-dimensional scaling" to create a two-dimensional diagram to show how close the CGM days are to each other. On the left is an example:



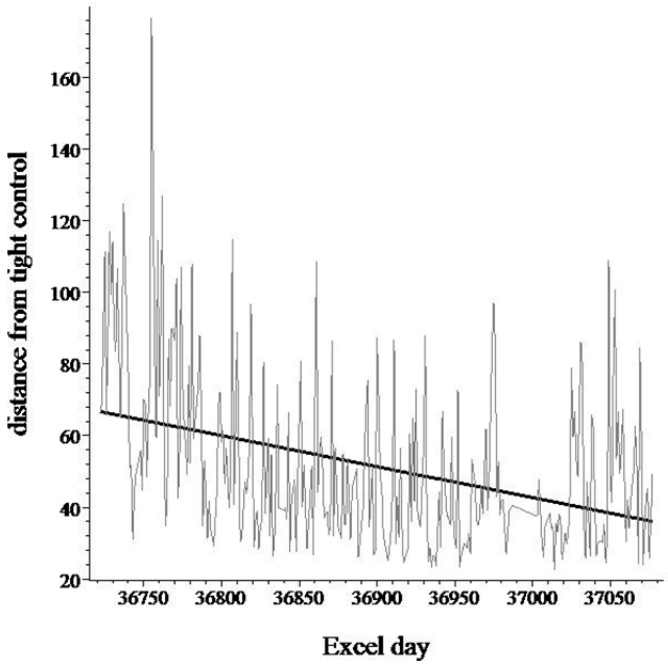


Here is how this diagram works: This diagram is for patient #374. Each black dot represents a day of CGM data. The grey circle in the middle is the "Region of Tight Control", which is defined based on the best CGM days in the database. (These are the days with little volatility and with glucose numbers close to what a non-diabetic has.) For patient #374, more than half of the days are in the "Region", and most of the rest are close, so this is good. There are a few days that are far from the "Region" and are not good days, such as the day at (-2, 4). On the right is a graph of the CGM numbers for that day – the patient had a bad evening.

On the next page is a diagram for patient #53, who is having more trouble with blood glucose management.



I also developed a second report to study how a patient's blood glucose management evolves over time. Here is an example of a patient who is a "quick study" and shows marked improvement over the six month period of that a CGM was being used.



What is going on in this diagram is that for each CGM day, we are graphing the "distance" from the CGM day to the center of the "Region of Tight Control", and then drawing a line-of-best-fit for the data. A declining line indicates that the tendency is for the distance to get smaller and hence for control to get better.

I also developed a second "similarity engine" that does not use wavelets. Instead, I sliced each day into 80-minute periods, and give each period a "grade" based on average blood glucose during that period, and to what extent the numbers are climbing or falling. Using these grades, I came up with a way to generate a score for each CGM days. The distribution of these scores appears to be on a bell curve.

These diagrams and scores provide ways to compare type-1 diabetics and to provide patients with ways to better understand their blood glucose dynamics. In the end, they capture the spirit of the question #2 above, and allowed me to answer question #1.

Writing and Dissemination

I anticipate producing two papers based on my sabbatical work. One of the papers would be the wavelet-based analysis, and it would be targeted to a journal read by mathematicians and mathematics students. In fact, at the end of December, I completed a solid draft of this paper, and I plan to submit a final version in February. The second paper would use the grades and scores, and it would be for an audience of diabetes educators and researchers. I will have this paper completed and submitted by July 2012.

During my sabbatical, I was invited to speak at the Honors College at Michigan State University about my work in progress. At that point of time (early November), I had completed the "engine" and had some preliminary results, but not the diagrams described above.

Also during my sabbatical, I wrote a two-part mathematically-themed article for the *Chronicle of Higher Education* web site, dealing with mathematics in a lawyer drama on television called *Suits*. (Google: Aboufadel Suits). In November, I learned that a short article on the mathematics of paint tinting that I submitted earlier in this year was accepted, and I spent a day revising that article. Then, in December, I learned that research done during Summer 2011 on the detection of potholes in data collected by smartphones was one of three solutions that were accepted as prize-winners by the InnoCentive company and the city of Boston. (Google: InnoCentive Boston pothole). I spent a few days doing some work to make our code publicly available.

Conferences, Travel, and Learning

I attended two conferences during my sabbatical. One was a conference in Grand Rapids – GRRCon – that dealt with computer security. While not at all directly connected to my sabbatical research, it was interesting to learn about new technology such as the "smart meters"

used to record electricity use, and methods being devised to protect such technology from hackers.

The second conference was a conference on large data sets that was held in October in Silicon Valley, California. The conference was hosted by NASA, and I met many researchers in the field of large data sets, mostly from other disciplines like computer science, and meteorology. In fact, I believe I was the only mathematician there. At the conference, I learned about a number of techniques to work with large data sets, and came up with several ideas to try out on CGM data. Most of these ideas did not pan out, but the work did help deepen my understanding of the dataset I was working with. I was also pleased how successful my background preparation was for my sabbatical, as I could follow nearly every talk at the conference, and I was already aware of many of the methods were applied.

There were also at least a dozen papers that I read closely either before or during my sabbatical. One, which discussed ways to sort out how much of success in a task is due to skill and how much is due to luck, drove much of my thinking during September and October, and was the basis of question #1 above. I did come up with a result that good blood glucose management appears to be connected to the education level of the caregiver (patient or parent), with the more education, the better, which suggested that there is a lot of skill involved.

Overall, I am quite satisfied with my learning during the sabbatical.

Implication for Future Scholarship

Since 2008, I have been working with different students on projects connecting the clustering of data with the applications of wavelets. The first project – looking at statistics from the NFL – did not really amount to much, but it was the type of failure that helped make the second project – my first foray into type-1 diabetes – a success. (The students won a mathematical biology prize and the article was published in a peer-reviewed journal.) There was something unsatisfying about that project, though – a sense that the mathematics could be better.

With this project, I feel that the mathematical methods I am using have been properly refined. This implication is that I can now use these methods with confidence in future research projects (with or without students). So, I anticipate a student-collaborated wavelet-based applied mathematics project in the near future (either summer 2012 or 2013) that utilizes the techniques from my sabbatical.

I will speak about my sabbatical results at the 2013 Joint Mathematics Meetings, will be held in San Diego.

Other Notes

In my proposal, I noted that I was going to apply for an external grant to help fund my research and travel. I applied to the Juvenile Diabetes Research Foundation in February 2011, and after several months of discussion back-and-forth with them, they declined to fund my $15,000 request. Part of the issue was that my research is mathematical rather than clinical and they didn't know how to fit what I was doing in to what they wanted to fund. The other was that they said I didn't ask for enough money, as they were looking to fund bigger projects. So, I do not have any external financial remuneration to report. Nevertheless, I hope that the researchers at JDRF are impressed with my results.