Journal of Diabetes Science and Technology
Volume 5, Issue 3, May 2011

SYMPOSIUM

Pediatric Diabetes Registries: When Baby Steps Are Not Enough

Chad Malone, M.D., and Gary D. Kennedy, M.B.A.

Abstract

Effective diabetes research relies on pattern recognition. Although information technology (IT) has been used to aid researchers in recognizing patterns, there are still barriers to effective data collection, analysis, and collaboration inherent in using outdated methods and technology designed to fulfill clinical, not research, purposes. This article discusses seven problems with current research and outlines a solution in which innovative IT can be harnessed to overcome each problem, resulting in better research outcomes.

New IT solutions on the market, such as meta-registries, are designed specifically to handle the complex data collection and analysis problems associated with diabetes research. A meta-registry with an ontology automatically harmonizes data from disparate sources, allowing researchers to devote their time to pattern recognition. With all essential data centralized and harmonized, researchers are also provided with a more complete view of each patient or research subject. When researchers can view and report across all data types at the same time, they are able to discover patterns and associations that are indistinguishable using traditional methodologies. This capability proves extremely beneficial, particularly for multifactorial disease research such as diabetes research.

J Diabetes Sci Technol 2011;5(3):571-578

Diabetes research is fundamentally a sophisticated exercise in pattern recognition.[1] This is a difficult endeavor, because there are complex interactions between multiple affected body systems to consider. Identifying patterns has become much easier with widespread adoption of information technology (IT) throughout health care. However, as IT has found its way into health care, it has become obvious that there are areas in which IT can make a huge difference, areas where it is useful but not "game changing," and there are a few areas where technology has brought minor improvement or even resulted in a step or two backward.

Information technology is more beneficial in solving some types of problem than others. This is true for all industries, including health care. An example from the automotive industry illustrates how IT is sometimes harnessed injudiciously to tackle minor problems *in lieu* of major issues. Andy Grove, the past chief executive officer of Intel, once taught a class where a sales executive entered the room and interrupted to announce that Intel had just received its first purchase order from Ford Motor Company. The sales executive wanted Grove to see it because this was a long-anticipated event at Intel and widely regarded as a harbinger of great things to come

Author Affiliation: RemedyMD, Sandy, Utah

Abbreviations: (HIPAA) Health Insurance Portability and Accountability Act, (IT) information technology, (RDR) registry data repository

Keywords: diabetes data collection and analysis software, diabetes research registry software, RemedyMD, tools and applications

Corresponding Author: Chad Malone, M.D., RemedyMD, 9350 South 150 East, Suite 850, Sandy, UT 84070; email address *cmalone@remedymd.com*

in the automotive segment. Grove studied the order for a few seconds and then, to the class's astonishment, threw it down in disgust. He then asked the following question: "When are the executives at Ford going to understand that microprocessors ought to be used for sophisticated applications like automatically adjusting the fuel/air ratio for high altitudes or sensing worn brake pads on the rear wheels so they could apply stronger pressure on the front brake? Using microprocessors for these trivial options like toggling between Centigrade temperatures and Fahrenheit temperatures is a waste of their money and a silly use of our innovations." In time, the automotive industry utilized IT to enable their vehicles to do previously impractical or impossible functions. All who are involved with funding, managing, or receiving health care are hopeful that this industry will find similar opportunities to transform itself through the application of IT.

The judicious use of IT can overcome the following types of research problems:

1. *Data problem*: Researchers are faced with a cacophony of data streaming in from many directions. In addition to new types of phenotypic data, researchers are now able to analyze molecular data, which creates entirely new challenges.[2] The heart of the problem is that pediatric diabetes researchers often have a clear vision of what they want to accomplish but are stuck on the tedious tasks of managing the underlying data and spending their valuable research time building interfaces to gain access to their existing data.

   Most researchers are aware that the most efficient place to ensure data quality is at the point of entry, so they try to build or acquire systems that have sophisticated tools to ensure accurate data collection. This is an effective strategy for collecting their own data, but pediatric diabetes research relies heavily on data collected from other sources, such as electronic medical record data, patient-reported longitudinal data, laboratory data, and, soon, molecular data. The data problem lies in finding a way in which data from other sources can be curated or at least manipulated into a form that integrates easily with the researcher's existing data. For many years, pediatric diabetes researchers have spent a significant amount of their research time figuring out how to aggregate, link, and harmonize data from numerous disparate sources.[3]

   In addition to wasting the researcher's time, the act of manually curating data presents other problems as well. When data are forced into a new structure without being properly linked to other data elements and harmonized by a standard ontology, then the data become unreliable and of questionable value for identifying patterns. Accurate data analysis relies on consistent and reliable data. Through integrated applications and tools, IT can provide researchers with the ability to aggregate, link, and harmonize data in a way that allows them to see an integrated 360° view of the research subject or patient.[4]

2. *Multivariate problem*: Researchers are familiar with the complex and tedious process of linear regressions. These calculations are extremely difficult when a single dependent variable is plotted against a known series of independent variables. However, in health care, the problem is even more complex. Often, it is necessary to analyze the effects of two or more dependent variables. When this is the case, any hope of a manual analysis is abandoned. Multivariate statistics require sophisticated analysis techniques that are only possible through the use of IT. The only alternative is to create a simplified model that tries to approximate the reality being modeled. The oft-quoted phrase, "Essentially, all models are wrong, but some are useful," is especially true in health care.[5]

3. *Contextual problem*: Current research typically analyzes data without factoring in their context. Information technology can allow researchers to track provenance, origin, and context through meta-data tags. Analyzing data and their context may lead to more useful clinical and research decisions. For example, we can consider the way that glucose levels fluctuate based on the amount, type, and timing of food intake before a test. A more complex case using the same measure, or even more involved hemoglobin A1c levels, would involve comparing glucose or hemoglobin A1c levels for two children over an extended period of time. However, each individual metabolizes food at a different rate, depending on his/her genetic characteristics. Tracking food intake, without the context of their genetic makeup, can be misleading. In another example, two children from different families are being tracked longitudinally to measure their quality-of-life scores. If their parent fills out

a weekly survey, it would be important to consider the context of the care provider. It is possible that the parent or care provider's mood will affect their survey answers and, subsequently, the quality-of-life scores for each child. In each of these two examples, the context of the data is as important as the data.

Any meaningful attempt to include context along with data requires sophisticated IT, because the numbers of nodes, levels, and branches that arise soon exceed the ability of humans to consider each possible path. Lacking sophisticated IT means that a researcher needs to make some *a priori* decisions to limit the number of paths they can pursue. Making "research triage" decisions that have been made by other researchers will generally result in analyzing the same paths through data that have already been examined. This scenario results in few, if any, new discoveries.

4. *Wrong tool/Right job problem*: While there are varied opinions on the state of clinical computing, there seems to be universal agreement that research computing is not as mature as clinical computing. Why are these researchers using the most primitive tools and applications? The short answer is that the vast majority of IT applications or tools used by researchers today were designed for the clinical side of medicine, not for research. The problems solved by an electronic medical record are very different from the challenges researchers face when they try to aggregate and report across hundreds of little databases that are in their own format and then include myriad spreadsheets that also serve as data storage systems.

   Electronic medical record systems exist to capture and store billing, coding, and event-based data. They serve as the primary data source for most patients at large institutions. Although they serve their purpose well, they are not designed to (nor can they) collect genomic or biospecimen or patient-reported data. They cannot be modified on the fly to accept new data elements that have suddenly become important to a researcher. Electronic medical record systems are often based on long dictation notes, in unstructured text, from physicians' observations on a specific patient. These unstructured notes are helpful if used in the way one might use a digital recorder. You talk into the recorder, and it will later play back exactly what was said. Unstructured notes are of very little value to researchers. It is possible to use artificial intelligence and machine learning devices to gain a close approximation of what took place in an encounter, but the data collected in free-form text is too abstract and imprecise to be useful in making inferences about a patient.

   There is no doubt that electronic medical record systems are important, but they are not effective tools for conducting research. For example, suppose you purchased an electric can opener. It works well if you want to open a can of soup for dinner. This machine is designed to open containers and facilitate the removal of its contents. One may assume that, given its stated purpose, the new electric can opener should be a fast and easy way to open a bottle of soda. Not only is it ineffective at opening that particular container, but it may also puncture the lid and spray the contents all over you. It should work because it was designed to open containers and remove the contents. Furthermore, the electric version is even faster than the old type of can opener, yet it does not open a bottle. It might open a can of soda, but you would likely spill the contents, and you might cut your lips when you drink from the can. There is nothing wrong with the electric can opener. It does what it is designed to do very well, yet it cannot accomplish other related tasks. We have experienced a similar issue in that we have been trying to use the wrong tool to solve a problem that it was not meant to solve and is therefore not capable of solving.

5. *Collaboration problem*: The National Institutes of Health, through its Clinical and Translational Science Awards, has made collaboration a part of every researcher's life. Not only is collaboration a requirement for a Clinical and Translational Science Award, but it has become a factor in disease research that is not funded by the National Institutes of Health.[6] Of course, the scarcity of patients has forced researchers of rare disease to collaborate with other centers just to find a meaningful number of patients. This article has already discussed some of the reasons why an improved IT infrastructure is necessary to conduct pediatric diabetes research inside a single institution. Obviously, the problems wrought by a central hub that must collaborate with 20 other institutions are greater than those faced by a single institution. With a slight twist to Metcalf's Law,[5] it is possible to create a "complexity corollary," which states, "The complexity of a research

network increases by the square of the number of collaborators." Information technology can provide a solution that is accessed by multiple institutions and that includes an ontology for harmonizing all data. This type of system facilitates constructive collaborative research efforts.

6. *Multifactorial disease problem*: Certain pediatric diseases and conditions are relatively atomic in the sense that their effects are isolated to one area of the body, such as Hirschsprung's disease. Other diseases such as diabetes can affect the heart, the kidneys, the liver, the limbs, the eyes, and, of course, the entire digestive system. These multifactorial diseases are much more difficult to research, because the ramifications affect so many body systems. For example, a pediatric patient with diabetes could be seen by a variety of specialists depending on the type and the severity of the disease. They may be part of multiple disease registries and multiple patient registries. In addition, a particular device provider may also track the patient. Yet there is no automated way for a pediatric diabetes researcher to conduct any sort of cross-disease research on the child. Information technology can provide a system where a complete view of a patient includes information from their endocrinologist, cardiologist, and nephrologist.

7. *Data security*: The security issues surrounding the use of patient data in registries has long been a debated issue. In the past, there have been challenges keeping that data secure. However, all registry software firms have adopted procedures as set down under the guidelines of the Health Insurance Portability and Accountability Act (HIPAA) of 1996 for securing patient-centered data. First, to ensure that this data is secure, the software must be HIPAA compliant using Secure Sockets Layer, which protects from unauthorized attempts to access patient data. This means that only authorized persons can gain entry into the system to create a data record, retrieve stored data, or modify patient data. The software system provides for different levels of authorization and keeps a log of every user's access and exit to provide traceability. Second, database encryption makes data more secure now with two forms of data protection. This encryption means that the data and information stored therein has been converted into a unique code, which cannot be deciphered by unauthorized persons. Finally, session timeouts are required under the new HIPAA compliancy rules. With these new data protection mandates, data are much safer today than they were even in 2006.

Information technology has produced a model in which all data from all sources are brought together in a flexible and harmonized way that becomes the center of your data universe. This is commonly called a registry,. sometimes referred to as a research data repository (RDR). While a meta-registry does fit most of the criteria for a RDR, it differs in the sense that it provides its own highly flexible data model so that it is not necessary to have a large IT group to manage the registry. Because the meta-registry establishes links between each registry, cross-disease research is not only possible, it is automatic. Obviously, it is a very difficult task to use a RDR to conduct cross-disease research between two widely different diseases. This meta-registry can be utilized in a way that enables a researcher to analyze the characteristics of a diabetes patient without knowing which data fields come from which source. Or it can be segmented to show a cohort of diabetes patients who fit a unique set of characteristics and all the information that pertains to them from every source in the data universe.[7]

The premise of a meta-registry is that a researcher can recognize trends and patterns in data if and only if all data are stored in a harmonized way. Once the aggregation, linking, and normalization take place, then researchers are able to look at data through a variety of lenses or portals into the data. Our experience has shown that, when researchers can view and report across all data types at the same time, they are able to discover patterns and associations that are indistinguishable using traditional methodologies.

A meta-registry is populated by one or more electronic medical record applications, as well as various other sources of clinical and research data. Patient or research subject information can also be loaded directly into the registry after passing through the harmonization layer. It is by design infinitely flexible. Because it stands alone, its capabilities and limitations are independent from the structure of any source system.

Depending on your perspective, a meta-registry that offers limitless flexibility, offers easy interfaces to a variety of external systems, and enables researchers to dramatically improve their productivity is either a research miracle or a spreadsheet with an ontology built on a relational database management system.[8]

Funding challenges, pressure to produce results, requirements to publish new research breakthroughs, and an ever-increasing public desire to be treated by the real innovators have caused the medical research community to move from a model that could be called the "whatever the clinicians are using" model to the "registry-centric" model. **Figure 1** presents an example of a registry-centric model. Notice that data are collected from widely disparate sources on the left of the diagram. It is passed to a meta-registry, essentially a registry of registries, and is made available to principal investigators and other research-oriented groups. Data coming into the meta-registry is often secondary-use clinical data, although many researchers are beginning to use a meta-registry as their primary data store. Notice that no data can be stored until they pass through the ontology to be harmonized with relevant standards such as International Classification of Diseases-9, Systematized Nomenclature of Medicine, RxNorm, Unified Medical Language System, Logical Observation Identifiers Names and Codes, and Gene Ontology. The process of using the computer to interpret and enforce rigid terminology standards is essential to producing consistent research findings. If there are tools and applications available with the meta-registry, they are also made available to the researcher. **Figure 1** depicts the important relationship between tools, applications, and the ontology. All tools such as electronic data capture must be under the control of the ontology in order to ensure that spurious data can not be captured through circumventing the ontology. Reporting, query, analysis, and data visualization tools must also be fully controlled by the ontology if a researcher wants to have confidence in the results of a report. **Figure 2** contains part of the electronic data capture screens for a meta-registry. Notice that it is specific to pediatric diabetes. This pediatric diabetes registry

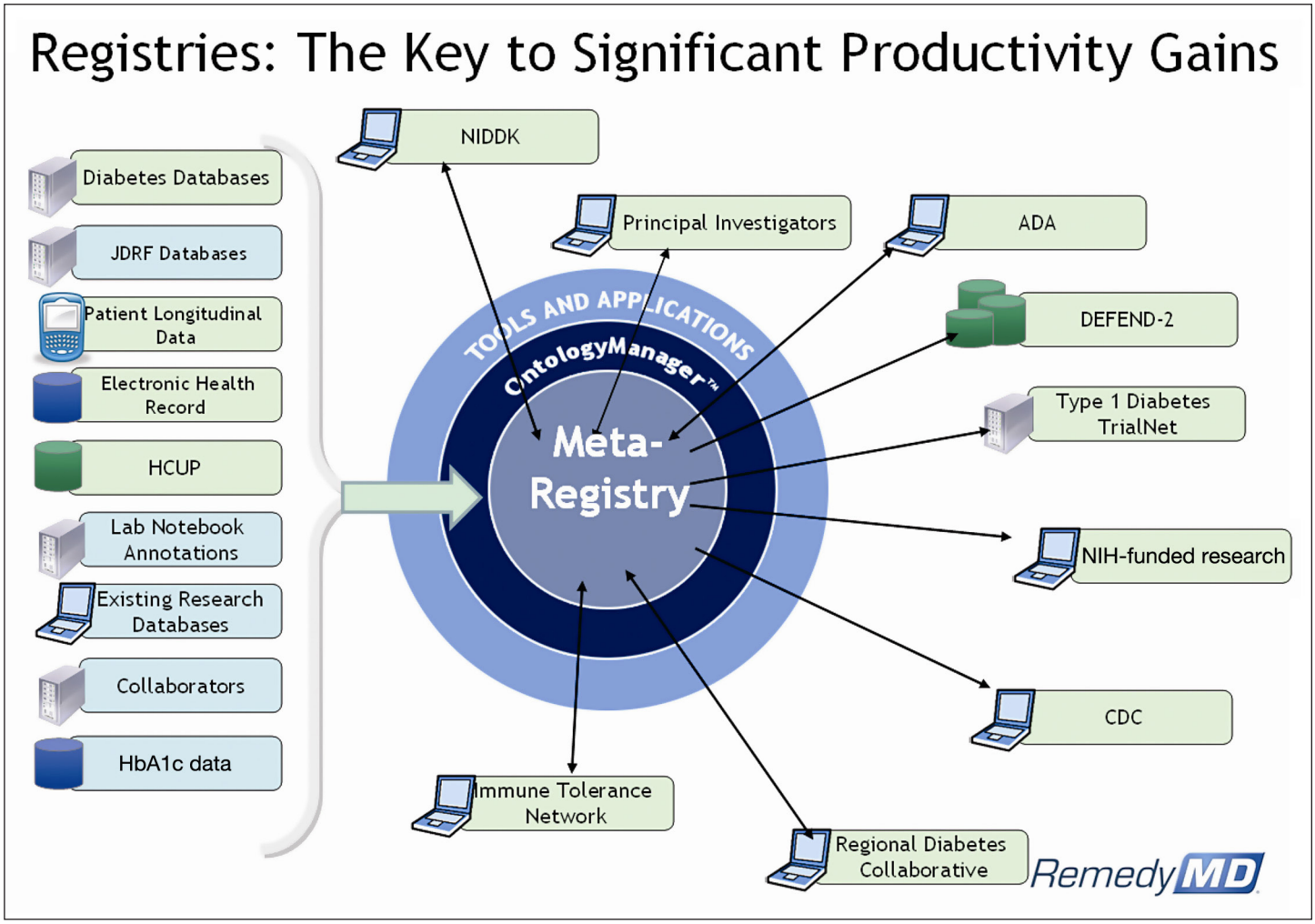

Figure 1. Representative of the optimum registry-centric model. NIDDK, National Institute of Diabetes and Digestive and Kidney Diseases; ADA, American Diabetes Association; JDRF, Juvenile Diabetes Research Foundation International; HCUP, Healthcare Cost and Utilization Project; HbA1c, hemoglobin A1c; NIH, National Institutes of Health; CDC, Centers for Disease Control and Prevention.

contains over 1500 fields that are specific to pediatric diabetes in addition to several hundred more that are common to all registries.

Figure 3 demonstrates an example of how it is possible to use data visualization tools to increase the strength of the information stored in the registry.

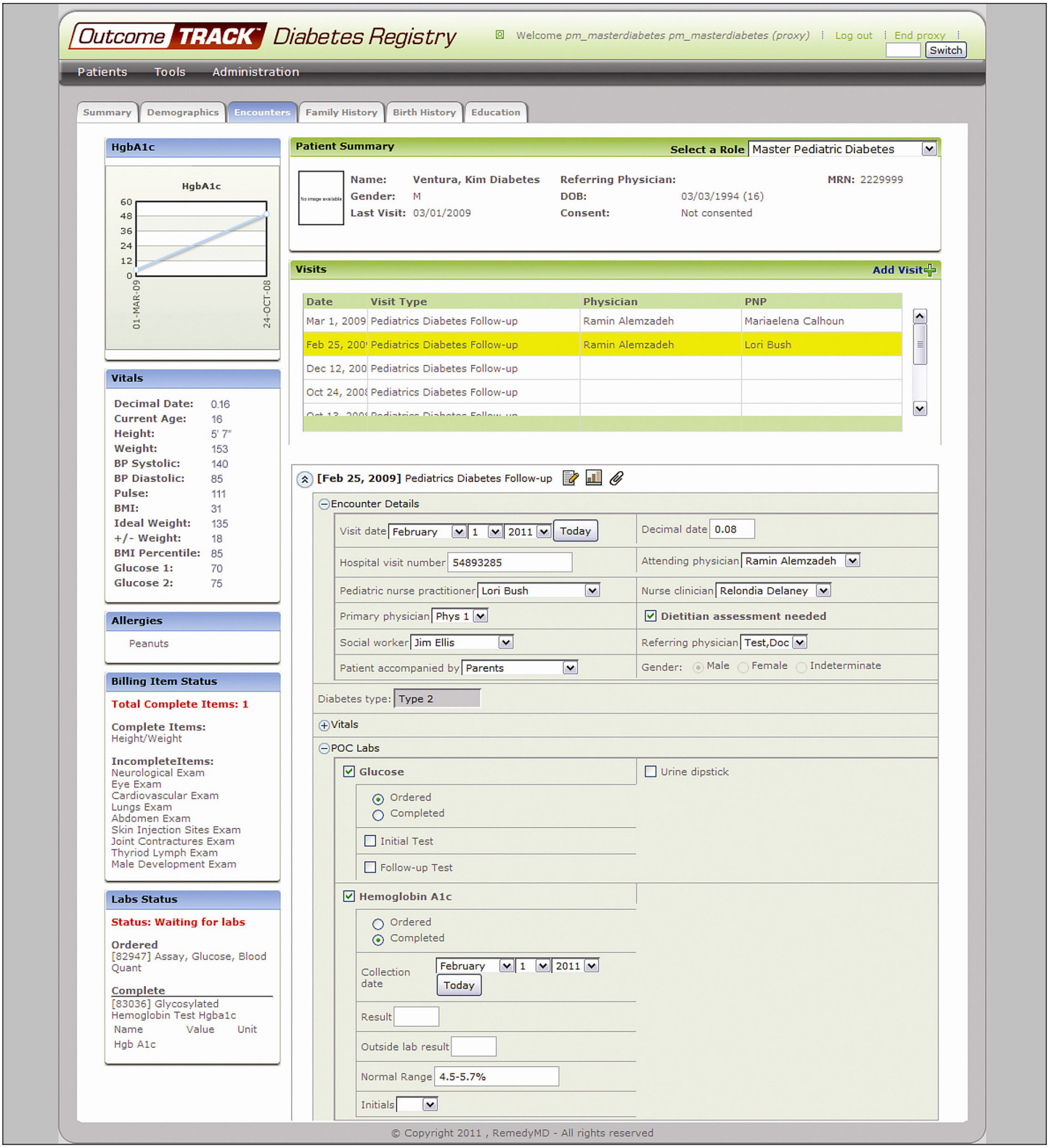

Figure 2. Example of electronic data capture screens for a meta-registry specific to pediatric diabetes.

The Mosaic Meta-Registry is being developed to be perfectly consistent with the diagram presented in **Figure 1**. Registries built on the Mosaic Meta-Registry are in production at over 100 sites. However, the current version of the product does not yet fully support all the features defined by this article. One of the most important features supported by robust research registries is the ability to capture event-based, temporal data such as individual patient encounter data over time. This type of capability is essential to be able to make every encounter a research encounter.

The most significant reason to implement a pediatric diabetes research registry is to realize significantly more effective research. If a meta-registry can handle seemingly mundane data management issues that are vexing researchers today and provide tools and disease-specific applications to manage research studies and experiments, then this will give researchers freedom and time to focus on meaningful innovation. Enlightened researchers have begun to realize that they can use IT to obtain a competitive advantage in an increasingly competitive research environment.

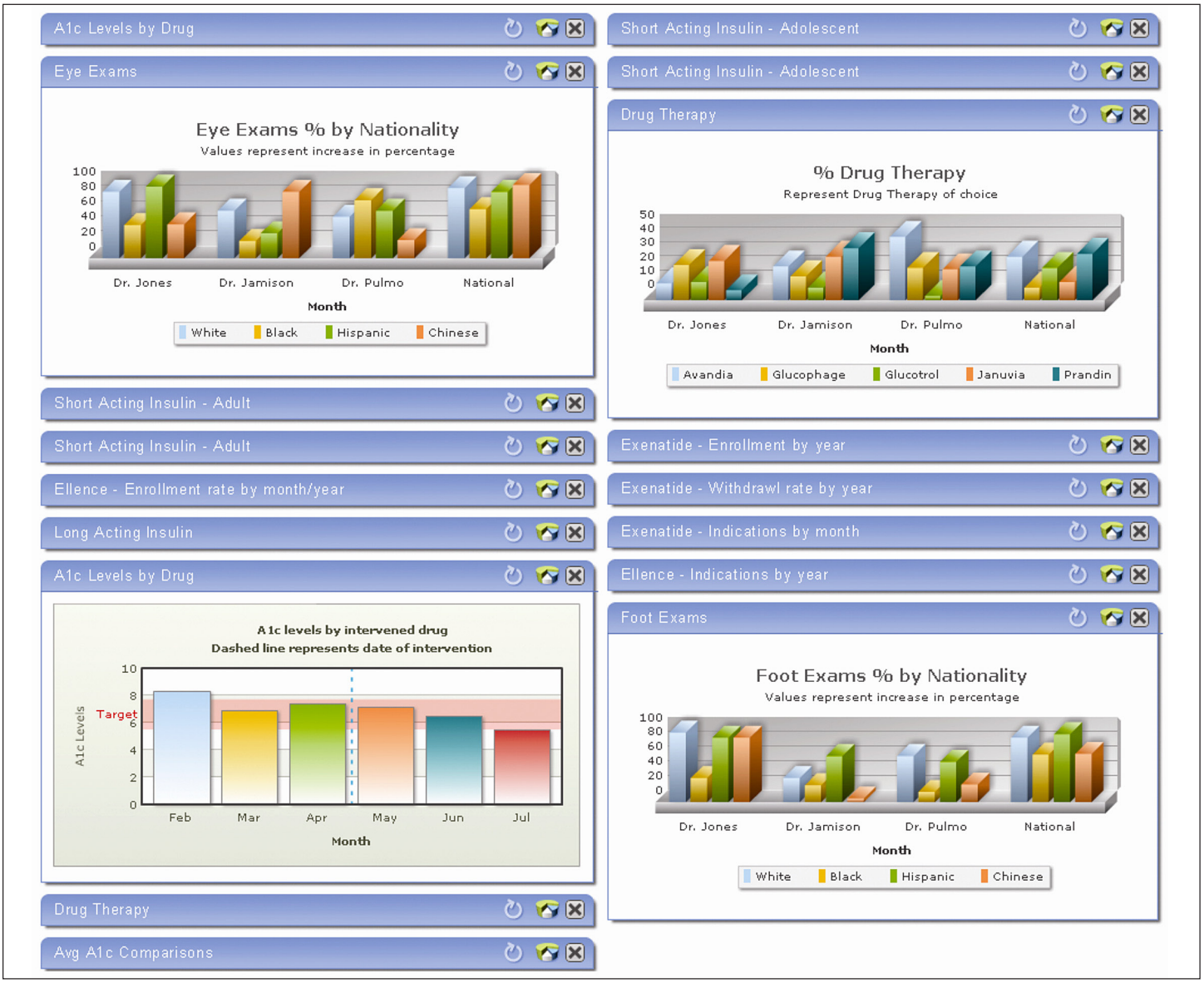

Figure 3. An introduction of data visualization tools allows researchers to recognize previously undiscovered patterns.

Disclosures:

Mr. Kennedy is the founder and chief executive officer of RemedyMD. Dr. Chad Malone is a paid staff member of RemedyMD.

References:

1. Dietert RR, Zelikoff JT. Identifying patterns of immune-related disease: use in disease prevention and management. World J Pediatr. 2010;6(2):111–8.
2. Vaught JB, Caboux E, Hainaut P. International efforts to develop biospecimen best practices. Cancer Epidemiol Biomarkers Prev. 2010;19(4):912–5.
3. Giannini C, Mohn A, Chiarelli F. Technology and the issue of cost/benefit in diabetes. Diabetes Metab Res Rev. 2009;25 Suppl 1:S34–44.
4. O'Connor M, Shankar R, Das A. An ontology-driven mediator for querying time-oriented biomedical data. Presented at: 19th IEEE International Symposium on Computer-Based Medical Systems, 2006. CBMS 2006.
5. HighContrast; Simeonov S. Metcalfe's Law: more misunderstood than wrong? *http://simeons.wordpress.com/2006/07/26/metcalfes-law-more-misunderstood-than-wrong/*. Accessed April 22, 2011.
6. National Center for Research Resources. Clinical and Translational Science Awards. *http://www.ncrr.nih.gov/clinical_research_resources/clinical_and_translational_science_awards/index.asp*. Accessed January 20, 2011.
7. Payne PR, Borlawsky TB, Kwok A, Dhaval R, Greaves AW. Ontology-anchored approaches to conceptual knowledge discovery in a multi-dimensional research data repository. Summit on Translat Bioinforma. 2008;2008:85–9.
8. Schweik CM, Stepanov A, Grove JM. The open research system: a web-based metadata and data repository for collaborative research. Comput Electron Agriculture. 2005;47(3):221–42.