Autism 360: The Development of an Online Database with Patient-entered Data

Sidney MacDonald Baker, MD

Abstract

This article describes a system and method for the automated presentation of data to and interaction with a computer-maintained database. The author invented, developed, and patented a novel information system, Medigenesis, and then created an adaptation, the web site Autism360.org, to serve the needs of individuals with autism spectrum disorder (ASD). The site creates an anonymous, multidimensional electronic record of profile items that captures a patient's individuality. An algorithm clusters users, permitting an exchange between individual and collective data.

The author provides a general overview of the system's design and presents an example of its power to reveal patterns. The article shows the system's potential for use in research in the context of an analysis of 79 482 profile items, comparing 1495 boys vs 336 girls with autism spectrum, in the context of groups who had either high- or low-normal body mass index (BMI).

Sidney MacDonald Baker, MD, an *IMCJ* associate editor, is the founder of Autism360.org, a nonprofit organization serving the global autism community to "let the data talk." Dr Baker is also an author and currently practices integrative medicine in Sag Harbor, New York.

Contemporary medicine lacks a scientific-notation system for recording and storing its most basic data; ie, the words that comprise the patient's open-ended flow of narrative describing signs, symptoms, life events, exposures, strengths, and other features that capture each person's unique story. Without such a system, the prevention and treatment of chronic illness remains dependent on a system in which those involved in medical practice, research, reimbursement, administration, and planning think about chronic illness within a false metaphor defining disease "entities"—fit only for the narrower limits of acute illnesses.

The definitions and boundaries in the fragmented landscapes of diseases and their medical specialties obscure common underlying physiologic and environmental factors. The very notion of chronic illnesses as entities tends to put the focus on treating the disease and on understanding the ways in which a patient fits its defining features. In opposition is a model that focuses on the ways in which medical practitioners might see patients as individuals with unique profiles defined by narrative and laboratory data. The author's previous paper presented[1] 4 fundamental principles of medicine as applied to the prevention and treatment of chronic illness and called for a need to capture the flow of individual medical narrative.

The current article describes the development of a novel information system that the author developed over time. With funding from The Moody's Foundation of New York, the author adapted that system to function as a web site, http://www. Autism360.org. The author hoped that the site could serve the interests of the global autism community, among which was the desire to have a means for capturing a detailed, accurate, structured portrait of each patient as an individual. Individuality is the fundamental principle of biology. For medical practitioners and caregivers, the disease's defining symptoms and its ominous name, autism spectrum disorder (ASD), can cast such a deep shadow that they lose sight of the individual patient's experiences.

An individual's depiction benefits from efforts that support careful listening to medical narrative. Careful listening involves feedback that says "Is this what I heard you say?," and any data-collection system that incorporates it must provide a method of recording conversations, a means of correcting errors, and approaches that allow its users to make sense of the data. Making sense begins with capturing the narrative's words and numbers as well as the resulting conversations that flesh out the meaning of the data. Information technology provides a method of storing that data to permit retrieval of individual and collective information. That record offers users a way to assemble the many fragments of vexing, curious, mysterious, and serious, past and present problems of an individual patient. The first beneficiaries of the database's storage of data are the patient and his or her parent(s), who become the owners of the anonymous web-based record, at no charge, and therefore, can store and have access to the patient's narrative. When a medical practitioner first pronounces the disease's name, caregivers often feel relief that "we know what it is." Relief's partner, grief, comes with their realization that the word *spectrum* removes the clarity that a diagnostic focus might provide. This realization magnifies their emotional upheaval.

As found in the author's database, the fragments of a child's individuality, including strengths and special skills, can provide a focus and a source of actionable options when presented in two ways: (1) as a tool for depicting and tracking the status of many features of the disease in response to diagnostic treatment trials and therapies over time, and (2) as a means for an interchange between a single individual's narrative and the experience of many other individuals. The database allows an in-depth analysis of the collective experience of its users because it captures the individuality of its members with detail, accuracy, and structure. Its data can be a rich resource for research, while at the same time the database meets the needs of patients and caregivers. The database's structured gathering of the fragments of information that make up individuals' medical narratives can allow researchers to integrate that data to serve healing. Enriching the imagination that practitioners need for tailoring treatment, the database clothes the bare bones of the diagnostic label with all the elements that restore a patient's individuality. Researchers ultimately may solve the mysteries of the collective phenomenon not by understanding the bones of autism spectrum disorder but by discovering the patterns that a pool of individuals' details can make manifest.

METHODS
Autism360's Evolution

The author's mentor, Shannon Brunjes, made a comment in 1969 that medical narrative data must be put into rows and columns to become useful.[2] The author based the dimensionality of his coding method for his research on that comment. The author's invention of this system involved recording every word of every narrative of every patient of a general medical practice, which focused on chronic illness, for more than a decade, beginning in 1978. Eventually a series of computers replaced the pencil, graph paper, and big eraser that the author used to produce a working model. The author conceptualized this model as an n-dimensional hyperspace (ie, a database integrating a complex, multivariable situation), in which the language of the narratives defined the database's limits and structure, while the database provided the best compromise between completeness and economy. The author did all the coding to maintain the single viewpoint that benefits classification systems.

The author's 1992 publication of a comparison of 2 groups of patients with unmet needs for magnesium discussed the general design of his coding method and revealed its capacity to identify otherwise invisible patterns in the collective representation and analysis of clinical data.[3] The general medical context for that study shows how the coding system supports the entire lexicon of medical description. The author based his data collection for the magnesium study on the computer interface that he used in his office, in preparation for and during interaction with patients.

The spatial or dimensional qualities of the system are particularly apt for representing data that describes individuals with autism, a disease that currently stands alone as a condition associated with the word "spectrum" in its common usage in both vernacular and medical language. The word implies that individuals with ASD occupy a conceptual space that is very different from patients whose disease is an entity. At the very least, that difference moves the author's paradigm from a point in space to a line or ribbon representing the continuum on which medical practitioners place individuals. It is the individual, not the disease, that takes a place on the continuum, which includes others whose conditions are more or less similar, and therefore, whose place in the spectrum is more or less distant from the individual in question.

On the one hand, the autism spectrum is spare in its so-called defining features. On the other hand, children in the spectrum particularly demand complexity and nuance in the descriptive terms that portray individuality, especially among the statistical outliers who often inspire discovery. Autism360's methods for acquiring data are novel to the field. Information technology's democratizing amenities, with its capacity to embody the principle that the data be coded at its source, was the inspiration for Autism360. Which is to say that coding should be applied as a word or number enters the system and not retrospectively. The author believed that information technology best served the patient's (or parent's) and the physician's interest in the accuracy, level of detail, and thoroughness of a record.

Entry of data into Autism360 is voluntary on the part of parents, who became aware of Autism360 by word of mouth, from media available through networks in the growing autism community, and through web searches. At the time of the initial login, the user must enter only the month and year of birth, gender, height, and weight for the individual with ASD, in response to specific questions. The site provides a web-based interface that controls other data that patients and parents originate by responding to an invitation to "tell us everything you wish that captures your individuality as described in serious, mysterious, vexing, curious, and even trivial details." Once a relationship—however anonymous—with its users is established with the initial creation of a record the website has the possibility to ask them individually or collectively to respond to questions motivated by research or clinical interests that may constitute notification regarding patterns that may inform the users' clinical options.

The database provides a feature that qualifies an individual's record to be matched to a cluster of like individuals in the database. Clustering is an amenity of the website that permits users to find treatment options based on the experience of others in their cluster. The author based the matching on dot-product proximity analysis. To be matched, each record has to consist of at least 15 profile items, which must include one strength or special skill. The author otherwise applies no eligibility test to users of Autism360, whom the author has encouraged to remain anonymous. Taking notice or making a record of a patient's strengths or special skills is not part of contemporary medical practice. Such a feature suits the purpose of a data system that has the intent of capturing individuality and that solicits positive information from parents and children who might otherwise be overwhelmed by a flood of failures and worries.

The author chose the number 15 as the cutoff for inclusion of data in clustering based on his analysis of records of his patients' entries on a section of a prior questionnaire for individuals with

chronic illness. That questionnaire began with the respondent's open-ended creation of a list of present problems that the patient considered to be vexing, mysterious, or serious. The author made such a questionnaire a part of an Internet-based system. Comparison of the capacity of free-form or questionnaire derived data revealed equivalent efficiencies for clustering data by proximity analysis within the system's multidimensional space.

The interface of Autism360 follows a methodology that the author developed for http://www.Medigenesis.com, which is currently inactive. The idea behind Autism360's coding system was the creation of a database that allows patients (with their parents) to find their places among a cluster of their neighbors on a continuum, which is not simply linear but multidimensional. Moreover, their places are the product of the granular details of their narrative and laboratory data and not a matter of an assignment that a professional makes, which is not a source but rather a product. In other words, the terrain makes the map, not the explorer. Obedience to that precept will result in all chronic illness becoming conceptualized as a multidimensional spectrum in which commonalities will replace distinctions as the basis for thinking about prevention and treatment. The data, not the doctor, becomes the expert in assigning proximity. Patients, parents, providers, and researchers all become the beneficiaries.

Autism360's database consists of records that parents can use that allows them: (1) to communicate their child's profile more effectively to professionals, (2) to maintain information about treatments and treatment responses, (3) to track changes, (4) to see actionable treatment options based on the experience of other parents with children in their child's cluster, and (5) to contribute to a pool of information from which other parents, practitioners, and researchers may draw.

Data Entry

Users' choices when they enter data enable the system to code data for immediate clustering and later research.

Onset severity. This parameter is the score assigned by the parent or patient reflecting the severity of the profile item at its onset, which a parent or patient gives. When a user first records a profile item, the system offers four choices: (1) completely incapacitating, (2) severe, (3) moderate, and (4) mild, encoded with the numbers 12, 9, 6, and 3 respectively. For follow-up tracking, users record score changes directly in the 0-12 range, with 0 representing absent.

Vernacular name. For each profile item, the user can choose the vernacular name for a symptom in a number of ways: (1) by entering text in a search field that uses a 3-letter word match, (2) by clicking on a word displayed on word cloud (Figure 1), or (3) by selecting an anatomical location on a sketch that is matched to the user's age and gender (Figure 2). A choice by any of these 3 means evokes 1 or 2 layers of drop-down menus filtering the options to final specificity. Clicking *Behavior* from the options shown in Figure 2 delivers a drop-down menu of which Figure 3 shows the top of an alpha-sort of options.

Data Coding

In inventing the system's classification of the elements of

Figure 1. Shows Word Cloud, which is one of three options through which users of Autism360 can select an area of concern in a 2- to 3- step sequence of clicks to select a profile item.
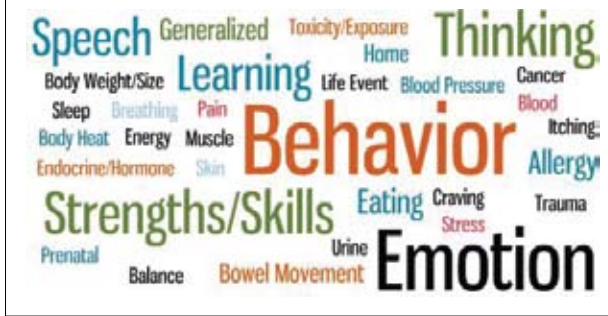
Figure 2. Shows a partial screen shot of Autism360.org's "Add Profile Item" interface showing the mannequin and the top of word cloud. On the bottom to the right is the top of the alpha sort of the menu that the user elicits by clicking *Behavior* on the word cloud. The long list of options presented by the choice of Behavior, with previously chosen items shown in red, shown with an overlaying dropdown fileter menu of which the first 5 are Activity, Aggression, Attention - Overfocus, and Attention-Underfocus, Clicking on any of these will drill down to specific vernacular options.

medical narrative, the author had to answer a recurring question when developing the system's lexicon of newly encoded words. That question was: "What is the literal meaning of the word, as free as possible from any implication of other considerations, such as causality?" If causality were to be implicit in the encoded word, as in the case of Herpes Simplex, the author would have to have accommodated a separate dimension (Etiology). By referring to descriptors as nonliteral, the author extends the concept of concrete location to qualify manifestations within systems such as behavior and emotion.

Table 1 lists the author's full set of previously defined dimensions available for encoding each element of a medical narrative.[4] It illustrates the complete list of dimensions available to describe a profile item, using an unrelated disease as an example. Shingles is presented as an example because it illustrates vernacular and technical synonyms and calls for descriptors in many

dimensions including Etiology. *System, Function,* and *Where* are obligatory dimensions for encoding the basic meaning of each profile item. The design criterion is invisible to the user and has no bearing on required fields of data entry.

As the online interface is used to enter patient information into Autism360's database, the system encodes the narrative details as intersections of 3 obligatory primary dimensions—*System, Function,* and *Where (Locale)*—and 1 quantitative dimension, *Severity,* for which a null value is acceptable. These dimensions provide descriptive terms for medical symptoms, life events, exposures, and other elements of medical narrative—as identified using a vernacular name—for a particular individual at a particular time and then quantify the details by temporal and other qualifiers as shown in Table 1.

	Table 1. Complete List of Dimensions Available to Describe a Profile Item		
	DIMENSION NAME	EXAMPLE	DIMENSION QUALITY
1	Formal name	Herpes zoster	Derived
2	Vernacular name	Shingles	Derived
3	ICD code	53.9	Derived
4	Modifier	Left	
5	Severity	Moderate	Quantitative
6	System	Skin	Obligatory
7	Function	Inflammation	Obligatory
8	Where	Chest	Obligatory
9	Onset	3 years ago	Time
10	Frequency	Constant	Time
11	Duration of episode	2 months	Time
12	Duration	2 months	Time
13	Precipitated by	Sun exposure	Factors
14	Aggravated by	:	Factors
15	Alleviated by avoiding	:	Factors
16	Alleviated by getting	Zovirax	Factors
17	Not alleviated by getting	Pain medicine	Factors
18	Etiology	Viral	Derived
19	Comment	:	Free text
20	Comment	:	Free text
21	Comment	:	Free text
22	Present date	20060925	Derived
23	Birth date	1/28/1937	Derived
24	Informant	(The patient)	Derived

System. Refers to the conventional body systems such as central nervous, respiratory, digestive, etc. In the present application, *System* also encompasses other elements, such as emotions, life events, and warmth.

Function. Refers to a literal interpretation for an encoded symptom, such as pain, itching, increase, decrease, various categories of abnormal, and descriptors such as strength.

Where. Indicates place, such as hands, eyes, or liver and includes whatever term captures the literal meaning of the patient's narrative. In the case of behavior, for example, the descriptors may include designations such as aggression, atten-

tion, and activity. These *Where* descriptors are equivalent to words that anatomically define the location of physical symptoms. The designations may enlarge or play upon the sense of occupation of a nonliteral space. That is to say, anatomical locations such as abdomen, finger, or even bowel movement are easily understood as having locality in the conventional—literal—sense. In the development of the coding system, the *Where* dimension received the exact or implied words that patients used to qualify the items of their narratives. The filtered list in Figure 2 gives an idea of the variety that characterized the *Where* dimension. See also Table 3.

Severity. Relates to quantification, encoded where applicable using a numerical scale as described below.

System, Function, and *Severity* are the dimensions that the author uses in the examples that follow. The graphics depict one entry selected from a single patient's record as the best method to explain the dimensions in the coding system.

Figure 1 shows groups that each represent profile items (ie, symptoms with their related descriptors) whose meaning falls within a given System-Function intersection. The figure shows the severity (vertical axis) for all profile items captured by the chosen entry into the system. For example, the figure shows the high level of severity of the functions *Abnormal, Decrease,* and *Increase,* for the system *Behavior* relative to the functions *Difficulty* and *Strength* under that same system.

Figure 3. *Systems* on the horizontal axis and *Functions* on the y axis define the intersections that carry the meaning of the profile items in the patient's record. Shown here are 28 out of 56 profile items in Autism360's record for a 12-year-old girl. The vertical axis shows the sum of the severity score at the time of onset of each item.
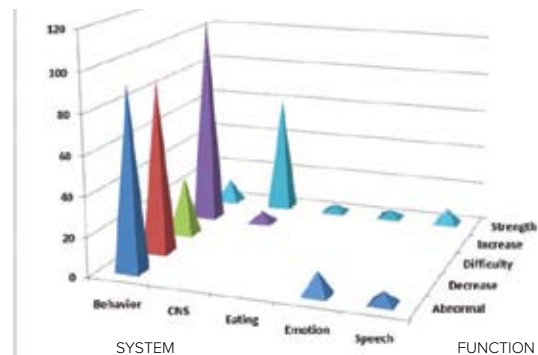

Figure 3 shows blank spaces at some intersections, and these blank spaces correspond to intersections for which the mother has not reported symptoms or for which combinations of *System* and *Function* are illogical (eg, *Pain/Speech*) and therefore null. The tallest peak in the graph, for example, stands at the intersection of *Behavior/Increase,* which embraces the symptoms that Table 2 shows, as recorded by the girl's mother.

The 6 behaviors in Table 2 all convey a sense of increase, as in "too much." The 4 locales—*Aggression, Attention, Activity,* and *Will*—exemplify the use of the *Where* dimension to encode

Table 2. Patient Symptoms in Vernacular Language With Associated Where Descriptor	
Behavior (System) / Increase (Function) Intersection	
Vernacular definition (medical symptoms in ordinary language)	Where
Violent outbursts	Aggression
OCD	Attention
Spend hours on pointless task	Attention
Stimming—throat clearing or coughing	Activity
Try to control others	Will
Stimming—scripting	Activity

nonliteral descriptions of the behaviors. These locales emerged from the descriptive language or narratives that gradually shaped the database's structure. The table shows the symptoms in vernacular language and the *Where* descriptor for those profile items at the *Behavior/Increase* intersection under the tallest peak in Figure 3.

In addition to information about the 3 primary dimensions, the database gives the researcher ready access to descriptors in each of 18 modifier dimensions, such as the vernacular and formal terms for the individual's symptoms, temporal modifiers (time of onset, frequency, duration of episodes, and overall duration), and source of information (mother, patient, doctor, lab, etc.) for the profile items in each peak of the graph.

Table 3 lists the severity scores (vertical axis) that produce the sum of 21 for the *Behavior/Difficulty* intersection peak in Figure 1, and the *Where* dimension fleshes out the meaning of the symptoms. The *Where* dimension exemplifies the latitude provided to anatomical location as an abstract or notional, rather than a physical, embodiment. The descriptor *Socialization,* like the *Where* descriptors in Table 2, are equivalent to words that anatomically define the location of physical symptoms. Finger-flicking stims are an example of an *Activity,* which captures the meaning of various stims that relate literally to an anatomic locale, which the Vernacular Name field identifies.

Participants

For the current example of the database's power to reveal patterns, the author's selection of participants differed substantially from that of a conventional project in which the researchers select participants based on strict diagnostic inclusion criteria,

such as the Autism Diagnostic Observation Schedule (ADOS) or any other validated test. Rather than contesting the normative value of such an approach, the author simply petitions for acceptance of a new Internet-based technology, http://www. Autism360.org, which gathers data using criteria that are substantially more open and which provides a means for revealing patterns that researchers might otherwise miss. By avoiding preconceptions about the object of his search, the author believed that he might detect signals in the data that would otherwise be denied to his perception.

Other fields of observation have promoted the value of such an approach. Very large data sets challenge researchers' capacity to predict or filter what is important; the author's approach instead relies on information technology to let the signal separate itself from the noise. Jim Gray[5] dubbed this approach *Fourth Paradigm Data Intensive Science.* Gray described[6] the evolution of scientific models, revealing four stages: (1) empirical—a thousand years ago, describing natural phenomena, (2) theoretical—the last few hundred years, using models and generalizations; (3) computational—the last few decades, simulating complex phenomena; and (4) eScience—today, facing a deluge of data that researchers capture with instruments or generate with simulators and then process using analytical and statistical software that reveal patterns.

The author selected 79 482 profile items of 1831 individuals between ages 2 and 19 from Autism360's database. The selection criteria for inclusion in this data analysis were the same as those that qualify an individual's record to be matched to a cluster of like individuals in the database. In choosing records, the author limited the choice of records only by excluding records that fail to meet the threshold of 15 profile items.
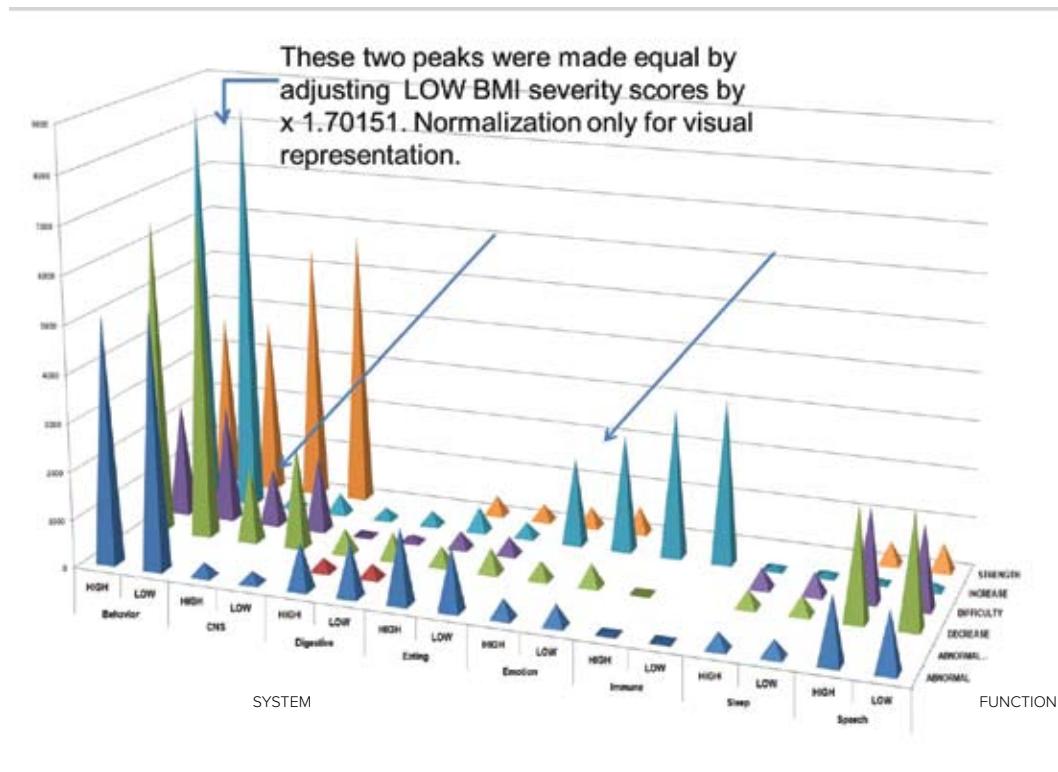
Data Analysis

The author filtered records containing 64 603 and 14 879 profile items of 1495 boys and 336 girls, respectively, and selected the profile items corresponding to individuals with BMI ranges >/= 5th to </= 25th percentile and >/= 75th to </=95th percentile. The author chose those segments of the normal distribution curve *a priori,* with an instinct to avoid extreme outliers. The author imported the entire body of data into Microsoft Excel, displaying it as intersections of the system's 38 systems and 43 functions. In graphics, those intersections display on the floor of the grid, and the vertical axis shows the sum of the onset severity of the profile items.

Table 3. *Behavior/Difficulty* Severity Score Data Depicted in Figure 1

Behavior (System) / Increase (Function) Intersection

Severity	Vernacular Name (profile items in ordinary language)	System	Function	Where
6	Peer relations problem	Behavior	Difficulty	Socialization
9	Problem engaging in conversations (poor joint attention)	Behavior	Difficulty	Socialization
6	Stimming—finger flicking	Behavior	Increase	Activity

Figure 4. Each peak in the illustration represents the *System/Function* intersection that codes the meaning of a subset of profile items that capture the medical narrative of 476 boys and 102 girls. These boys and girls, between the ages of 2 and 19 years, represent high-normal (75th to 95th) and low-normal (5th to 25th) Percentile BMI of Autism360's reported members. The author adjusted the sum of the symptom severity of the less numerous low-normal BMI group for the sole purpose of providing a visual anchor. Multiplication of that group's sum by a factor (1.70151) brings the tallest peaks at the *Behavior/Increase* intersection to exact equality with the peak (8562) of the high-normal BMI group.

The figure illustrates 8 *Systems* (Behavior, CNS, Digestive, Eating, Emotion, Immune, Sleep and Speech) and 6 *Functions* (Abnormal, Abnormal odor, Decrease, Difficulty, Increase, and Strength) among the 38 *Systems* and 43 *Functions* in the original data. Except as it applies to *Strength*, such as very good memory or math skills, higher scoring for profile items provides a metric for a negative or undesirable quality implied by the function. The arrows indicate eye-catching peaks that are discussed in the text.

These two peaks were made equal by adjusting LOW BMI severity scores by x 1.70151. Normalization only for visual representation.

RESULTS
Data Analysis

Among the data selected, the author selected certain System-Function intersections on the basis of being: (1) heavily populated, (2) suited to convey a visual compromise between economy and completeness, (3) representative of certain data of special interest, and (4) strengths.

Figure 4 illustrates these intersections, and it reflects the template for illustrating both the BMI- and gender-comparison data for the current analysis: (1) *Systems:* Behavior, Central Nervous System (CNS), Digestive, Eating, Emotion, Immune, Sleep, and Speech; and (2) *Functions:* Abnormal, Abnormal odor, Decrease, Difficulty, Increase, Strength.

As is visually apparent in Figure 4, certain intersections are blank because they are logically invalid. The function abnormal odor applies only to the digestive system among the systems illustrated. The author chose it to illustrate an example of logical invalidity when applied to most other systems and anatomical locations. The author also chose it because the microbiome

(microorganisms that are found in association with both healthy and diseased humans) has emerged as a focus for research in autism[7] as well as in health in general.[8]

For the purpose of data visualization, the author adjusted the BMI data by multiplying the severity scores of the low-BMI group by 1.70151, which created an exact equality between the 2 highest peaks in the illustration to provide a visual anchor. The author made this decision to make the first analytic step, visual inspection, easier, but he did not intend the adjustment as a statistical operation. The author intended the graphic representation of the data to give the observer a fast and keen appreciation of the contrasts and similarities among the profile items of children who differ only in the fact that they have a high-normal versus a low-normal BMI. For the purpose of this analysis, the author intentionally avoided children with BMIs below the 5th or above the 95th percentile, wanting more robust data in which the analysis could make clear distinctions based on the anticipated complexity of data at the low and high range of a continuum but could avoid data that describes individuals who are extreme outliers.

Body Mass Index (BMI)

Graphic representation of the data reveals a pattern of differences between males and females and between groups with high-normal and low-normal body mass index (BMI). As a group, females had a higher incidence of disabilities of the central nervous system, of emotional symptoms, and of immune sensitivities. The author found a very similar pattern of differences in children grouped by BMI, with the low-normal group resembling the females. Apart from those 3 differences, the males and females and the 2 BMI groups generally showed strikingly similar patterns of profile items, which demonstrates the capacity of Autism360's use of information technology to serve as a *macroscope*, permitting a grasp of otherwise invisible relationships.

Figure 4 and Table 4 show BMI comparisons, illustrating the overall similarity of the high- and low-normal BMI children. The arrows point to areas of conspicuous dissimilarity: *CNS/Decrease, CNS/Difficulty,* and *Emotion/Increase,* in which anxiety is a typical profile item. The distinction between *Decrease* and *Difficulty,* as applied to the central nervous system, comes from the descriptive vernacular, which uses the word *difficulty* to describe a particular cognitive or related activity. The *Eating/Abnormal* and *Eating/Increase* peaks are reassuringly consistent within the context of the data. Researchers have not reported a comparison of children with high-normal and low-normal BMI previously.

Table 4. Visual Comparison of a Subset of High-normal with Low-normal BMI Members of Autism360

	Count	Total Profile Items	Average Profile Items per child
HIGH BMI Boys	303	13 120	43.3
LOW BMI Boys	173	8044	46.5
HIGH BMI Girls	69	2468	35.8
LOW BMI Girls	30	1453	48.4
Total	575	25 085	

Gender

Figure 5 shows the same intersections as for the BMI comparisons and reveals a striking similarity to them, with the addition of the *Immune/Increase* intersection, which overwhelmingly represents profile items describing allergies or sensitivities to foods, medications, and inhalants. The figure for an adjustment of 5.2017 represents neither the ratio of the boy:girl profile items (4.341891) nor the ratio of the total number of boys to girls (4.449405). It simply represents the number required to bring the girls' *Behavior/Increase* count to equality with the boys for the sole purpose of providing a visual anchor—the tallest peak in the graphic—for the observer.

DISCUSSION
The Research Model

Autism360 follows the traditional clinical imperative of asking for a free flow of information from parents and patients.

The system gives that data a structure that is inherent in the literal meaning of the words. That traditional imperative of open acceptance of everything that the patient or parent wishes to tell, when taken to its logical extreme, says that the medical enterprise, the care of the patient, will benefit from accepting and respecting the entire flow of data that the patient can provide. The expectation of fulfillment of that potential is extravagant if researchers have only the doctor's traditional tools for transcribing, remembering, and considering the narrative of the patient(s), individually and collectively. Information technology permits our expectations to soar. If researchers respect the need to take as good care of the data as they do of the patient, then they will insist on effective means for acquiring, storing, and reporting accurate, detailed, and structured information: (1) by recording every detail of the narrative in real time from patients and parents, not clinicians, with provision for physicians' input and collaboration, (2) by storing the data electronically to permit individual and collective retrieval while protecting the anonymity and privacy of the patient, and (3) by analyzing the data according to the principles of what has recently been termed *Fourth Paradigm Data Intensive Science*[9] revealing answers to questions that researchers previously didn't know to ask.

Table 5. Visual Comparison of Girls and Boys Who Are Members of Autism360

|	Count	Total Profile Items (PIs)	Average PIs per child
Boys	1495	64 603	43.2
Girls	336	14 879	44.3
Total	1831	79 482	

This overall approach to medical information: (1) is patient-centered, supporting the patient's stake in telling and confirming the accuracy of and owning every aspect of the story he or she deems important, whether vexing, curious, or even trivial; it is not diagnostic in the sense that it intends to follow branches in the information tree to a terminal diagnostic leaf; (2) uses information technology to automate flow of information up from its original source, the patient, as a complement to automating the traditional habits and structures of top-down medical information from the expert opinions of physicians and the medical literature; (3) has an intention to "let the data talk" rather than use a hypothesis-driven model to drive research.

The results demonstrate the methodology's capacity to reveal patterns that are otherwise invisible to the naked eye and to give answers to questions that researchers would not otherwise know to ask. In the instance of the findings discussed in this article, such a question might be, "Do children with a high-normal BMI as a group resemble in certain kinds of symptoms those children who are male?

Autism, Gender, and Body Mass Index.

Examining every aspect of gender differences may illuminate the causes of autism. Informal polling of experienced clinicians

Figure 5. Each peak in the illustration represents the *System/Function* intersection that codes the meaning of a subset of profile items that capture the medical narrative of Autism360's members between the ages of 2 and 19 years. The vertical axis shows the weighted sum of symptom severity at time of onset. The author normalized the weighted sum of symptom severity of the less numerous girls' group for the sole purpose of providing a visual anchor. Multiplication by a factor (5.2017) brings the girls' tallest peak at the *Behavior/Increase* intersection to exact equality with the weighted sum (8562) of the boys' group. The arrows indicate eye-catching peaks that the text discusses.
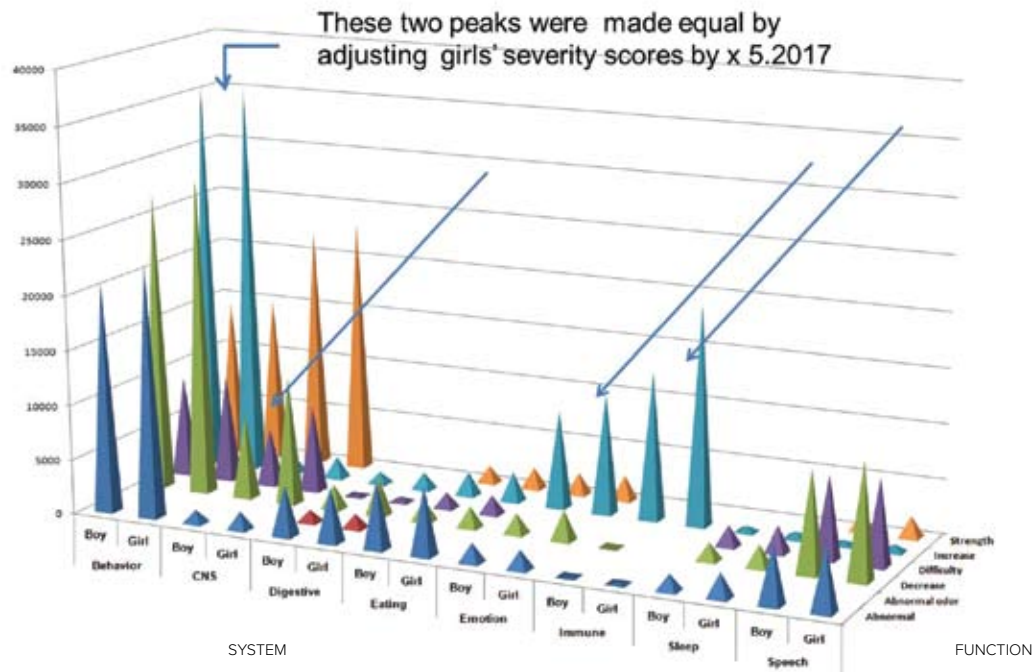
revealed a near universal agreement that autistic girls are very different from autistic boys. None of the clinicians, however, could specify with any precision or consensus what those differences are, apart from observations that girls are collectively less responsive to treatment, and therefore, in a sense harder to get both in and out of autism.

Research detailing gender differences in autism has been limited. "It's the white elephant in the room," says Marisela Huerta, clinical psychologist at the University of Michigan's Autism and Communication Disorders Center. "While many people speculate about the sex differences and a number of anecdotal stories exist about girls being clinically different or a bit harder to diagnose," Huerta says, "researchers haven't done a lot of empirical work on this topic."[10]

Body mass index is a conspicuous attribute of individuals that correlates with health status and fitness. In children, BMI is predictive of adult obesity, but researchers have published little correlating health status, fitness, and disease risk with BMI in children. A search of the literature reveals no published correlations of BMI with symptoms, strengths, or laboratory data in children with ASD. The published literature refers to the protective effects of estrogens and the susceptibility associated with androgen, but researchers have performed a paucity of comparative studies that offer objective measures of the differences with clinical details that might provide clues to answer the question "What does it mean that more boys than girls have the autism spectrum."

CONCLUSION

Presented in this article is the discovery of novel findings from the patterns of profile items in unprecedented detail that describe elements of the clinical narrative of nearly two thousand children representative of the autism spectrum. The very design of the system that enabled the acquisition of the anonymous data precluded obedience to the traditional requirement that each subject of the data be vetted as bona fide by standardized testing. By accepting that imperfections may find their way into the data, researchers may benefit from the strong voice of that data when orchestrated in the form of the visual analysis presented in the figures. The data then become the springboard for rigorous analysis of the correlations suggested by the data's patterns. The author does not include statistical metrics for the differences revealed in the illustrated patterns in this report but will present them in a forthcoming article. Those details would be misplaced in this overview because it focuses on patterns that precede measurement and raise questions for further measurement.

The analysis of gender and BMI identifies a close resemblance between low BMI children and girls and high BMI children and boys. The meaning of these gender differences and of the correlation of the symptom-pattern differences to gender are beyond the scope of this article. A glance at the position of the arrows in Figures 4 and 5 suffices to engender surprise. "Who would have thought?" was the author's first reaction to his confirmation that these close similarities between groups defined by gender and

low or high-normal BMI did not result from an analytical error or artifact but must derive from nature. A review of the relevant literature and detailed analysis of the findings illustrated graphically in this paper are forthcoming. The intent of this article is to present a full description of the information technology upon which the author based the findings and to put that technology in the following context.

The late 1960's saw the emergence of information technology's uses to computerize medicine. Since that era applications dealing with medical records have tended to automate the status quo, capturing text and delivering reports that deal with the coding of billing and diagnostic information. Hospitalized patients have benefited from laboratory reports that automate the presentation of flow sheets. The medical community has not exploited the capacity of information technology to aggregate and display patterns of clinical data generated during the care of chronically ill patients in an outpatient setting. The community has not developed the possibility of engaging patients in the creation, correction, and maintenance of the record of their own narratives. Researchers have paid little attention to revision of the antiquated infrastructure embodied in medical language that practitioners developed during the 19th century. Instead, communication about chronic illness continued to appropriate the paradigms of acute illness in which practitioners could ignore logical distinctions between the name and causes of trauma and acute infections without practical peril.

Nowadays medical practitioners suffer from the logical flaw encountered when a patient's behavior is attributed to autism; when a patient accepts the doctor's opinion that depression is the cause, not merely the name, of her sadness; or when practitioners decide that they should engage in a war on cancer. That flaw is rectified when ways of speaking, thinking, and acting acknowledge chronic illness as a complex interaction between the organism and its environment and not as the result of the attack of a disease entity. Medical practitioners can embrace the complexity of the interaction between organism and environment when they organize information that describes the results of that interaction—signs, symptoms, life events, exposures, lab tests, and results of treatment—with a focus on the individual and with a clear understanding that the individuality of each living organism is a primary principle of biology.

The relevance of that principle to the presentation in this article of the technology and results of a novel system for acquiring, storing, and analyzing clinical data, lies in the focus of that system on capturing individuality. The design of a system that supports an exchange between individual and collective data embodies a further principle, which demands that taking good care of individuals requires taking good care of their data and vice versa. The system acquires those data in a process that trusts the individual and collective intelligence of the user to the extent of shifting from emphasis on the traditional assumption that the doctor knows what questions to ask to the traditional imperative of recording the patient's complaints in his or her own words. The system described in this article places that imperative in a technological setting that permits the capture and analysis of details previously by-passed in the clinical exercise of naming the patient's problem. That exercise encounters traps of circularity in the path that begins with "we know what you've got." The implication that the name calls forth the treatment is an assumption underlying evidence-based medicine. Its value need not be dismissed by a shift to a focus on the data describing the individual in conjunction with the data defining the disease. The value of the author's presented system is that it allows the data describing individuals to redefine the medical community's concept of disease by its capacity to detect patterns not otherwise visible to the unaided eye.

Acknowledgement

References
1. Baker SM. Principle-based medicine. *Int Med Clin J.* 2011;10(5):22-33.
2. Brunjes S. An anamnestic matrix toward a medical language. *Comput Biomed Res.* 1971;4(6):571-584.
3. Baker SM. Magnesium in primary care and preventive medicine: clinical correlation of magnesium loading studies. *Magnes Trace Elem.* 1991-1992;10(2-4):251-262.
4. Baker SM, inventor. System and method for the automated presentation of health data to, and its interaction with, a computer maintained database, to generate information regarding possible remedies, therapies, problem solutions and beneficial practices, to improve user health. US Patent 7,676,384. September 15, 2000.
5. No author listed. Jim Gray summary home page. Microsoft Research. http://research.microsoft.com/en-us/um/people/gray/. Accessed April 6, 2012.
6. Hey T, Tansley S, Tolle K. Jim Gray on eScience: a transformed scientific method. Microsoft Research. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf. Accessed April 6, 2012.
7. Finegold SM, Dowd SE, Gontcharova V, et al. Pyrosequencing study of fecal microflora of autistic and control children. *Anaerobe.* 2010;16(4):444-453.
8. Nicholson JK, Holmes E, Wilson, ID. Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol.* 2005;3(5):431-438.
9. Hey T, Tansley S, Tolle K, eds. *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Redmond, WA: Microsoft Research; 2009.
10. No authors listed. Autism in girls: are the right questions being asked? Stone Hearth Newsletters. http://www.stonehearthnewsletters.com/autism-in-girls-are-the-right-questions-being-asked/autism/. Accessed April 6, 2012.