

Paper 166-2010

Analytics in the Converging Health and Life Sciences Industries: The Catalyst for the Next Generation of Healthcare

Matt Gross, SAS, Cary, NC

ABSTRACT

Learn how analytics serves as a catalyst for both healthcare reform and the foundation for the next generation of healthcare. Understand the hurdles and opportunities presented as health information converges from the health and life sciences industries and presents new business opportunities for each industry and improved quality of care for the healthcare consumer.

INTRODUCTION

Analytics in the health and life sciences industries has been sporadic, disjointed and hampered by numerous technical, political and financial obstacles. Many of these challenges are currently being addressed providing opportunities across the health and healthcare spectrum. This paper describes the evolution of analytics in healthcare and how a new generation of advanced analytics in the converging health and life sciences industries can fundamentally change the business and practice of healthcare. A perspective on how these advanced analytics are changing the environment from reactive to proactive, from isolated to comprehensive and from general to personalized. By focusing on creating an environment that enables a better and more comprehensive understanding of the individual, the condition, and the treatments, all participants of the healthcare system can more effectively and efficiently manage their health and focus on staying well rather than getting better.

HEALTH VERSUS HEALTHCARE

With so much emphasis recently on healthcare, it is important to note the difference between health and healthcare. Most people progress back and forth through various states of health. The majority of the population feel and, in fact, are well. However, many of these people engage in a lifestyle that results in their being at risk for certain events or conditions. When these behaviors persist, people will often contract a chronic condition, such as diabetes, that might need to be treated and managed for the rest of their life. Eventually, a person's health deteriorates to such an extent that it is no longer likely that the person's health can be beneficially managed and only palliative care can be offered. The diagram below shows the healthcare spectrum. While people might fluctuate between bands of the spectrum at times, in general most individuals will progress from well state through risks of conditions to being diagnosed with a condition and ultimately addressing the end-of-life phase. It is important to your health to understand exactly where you are in the health spectrum and to understand what your options are and choose a care plan specifically focused on providing you with the best outcome based on your personal preferences.

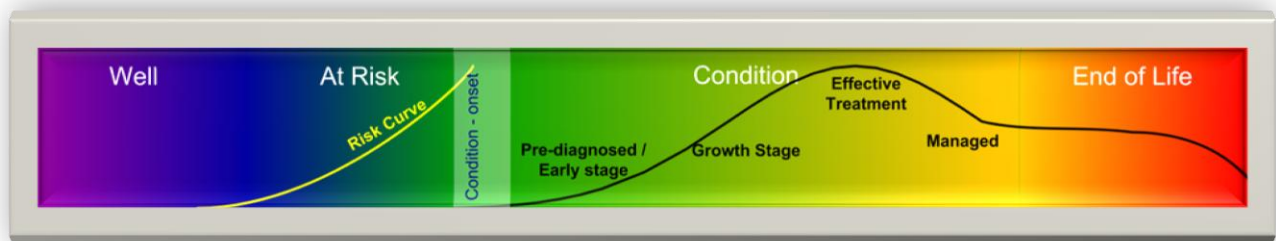


Figure 1 The Health Spectrum¹

Periodically, all people get ill or have an acute event such as appendicitis or a broken arm. These events can occur at any stage in the health spectrum and are usually treated as isolated events. However, each of these events can have

¹ Matt Gross, from a presentation on The Consumer-Focused Health Information Model, Helix Innovations, LLC, 2005

ripple effects into the individual's overall health in either a direct manner, such as the initiation of some sort of medication, or in an indirect manner, such as limiting a person's mobility, which can lead to exercise and nutrition issues. The figure below illustrates how acute or episodic events intersect with an individual's progression through the health spectrum. The cycles on the pink cylinders represent individual episodes and their event cycle. This cycle is discussed in the next section.

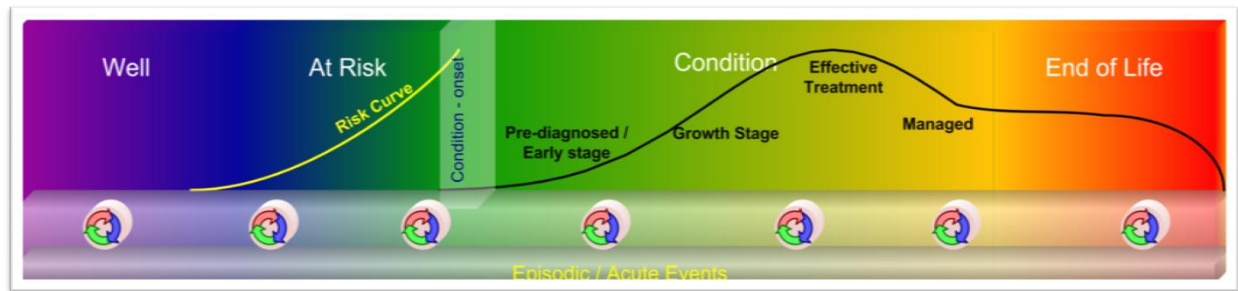


Figure 2 Acute events or episodes can occur at any point throughout the health spectrum

THE TRADITIONAL FOCUS OF HEALTHCARE

The broad purpose of healthcare has been defined by the goal of “Getting the right treatment to the right person at the right time in the right way.” However, I would argue that the main purpose of healthcare today can better be described as “How can we get this person better now?” The main difference between these two approaches is how much individual effort is put into the evaluation and care plan. Regardless of which version you believe is true, it is irrefutable that with very few exceptions, today's healthcare is reactive and focused almost exclusively on interacting with patients only once they become patients—in other words, once they've become sick. Only once a person is determined to be ill, is the healthcare system engaged with the goal of getting them to a better health state, whatever that might be. To do this, a simple four-step process has traditionally been used:



Figure 3 The traditional health event process

At each of the steps, a series of questions are asked to help guide the physician or medical provider determine the next step.

1. Assess the patient
 - a. What symptoms does the patient have?
 - b. What is their medical history?
 - c. What factors might have impacted or precipitated this event?
 - d. What are this patient's demographics?
2. Diagnose the condition
 - a. What is the likely reason for the symptoms?
 - b. What do these symptoms typically indicate?
 - c. What is the likely diagnosis for a person with similar demographics and history to the individual?

3. Prescribe a treatment or intervention
 - a. What are the available treatment options?
 - b. What has worked before?
 - c. What is likely to work best for this patient?
4. Follow-up
 - a. Is the patient better, worse or the same as before?
 - b. Do I need to alter treatment?

As you can see, there are a variety of data points to be collected and evaluated at each phase of the cycle. This data collection and evaluation is often done in almost complete isolation and is up to the diligence and competence of the individual physician and the patient. Data is collected at the physician's office, usually through a manual process where the patient fills out the ubiquitous clipboard questionnaire along with a brief conversation with the nurse and an even briefer conversation with the physician. That data collection process is then processed by the physician based on his or her personal experiences and whatever additional research he or she is able to do in order to come up with a diagnosis. After the diagnosis, based again on the individual physician's personal experiences and subsequent research, a care plan, treatment, or some other intervention is prescribed to the patient. The patients are left to their own devices to implement such recommendations. A follow-up usually occurs only if the patient self-assesses that they are not getting better.

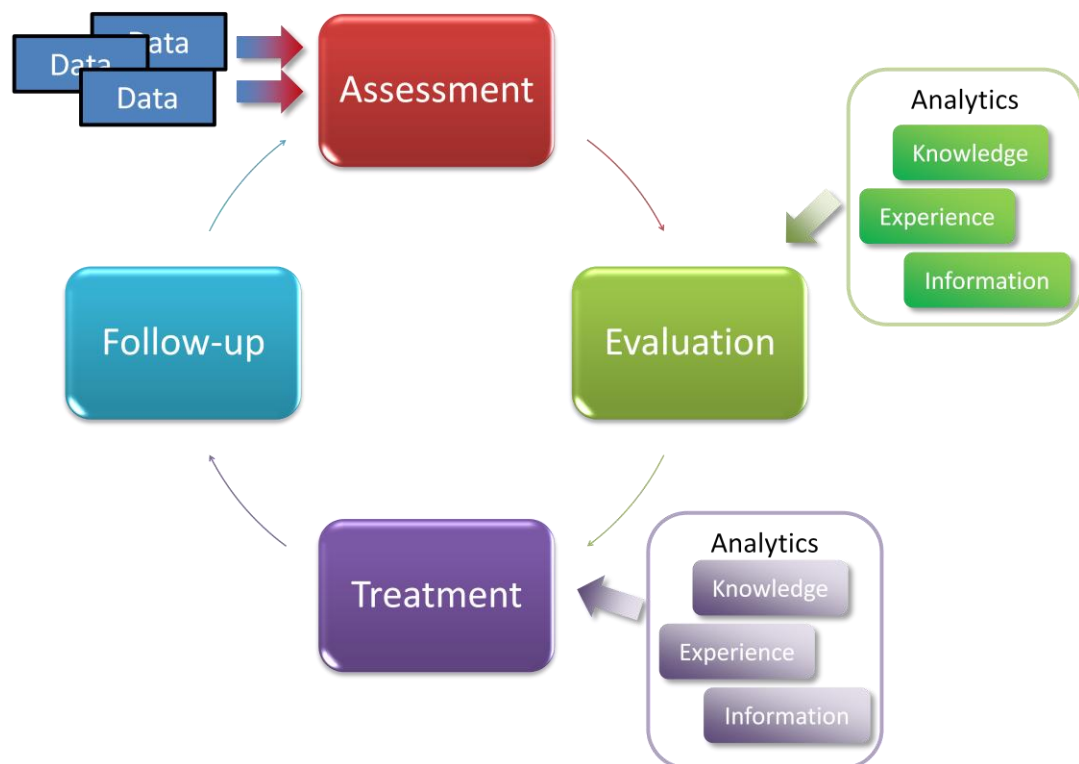


Figure 4 Data and Analytics in the Traditional Healthcare Visit

HEALTH INFORMATION AND TECHNOLOGY – A LONG, PAINFUL COURTSHIP

In the previous example, all information used for in analyzing the assessment, evaluation, and treatment was provided within the confines of that specific physician-patient interaction. Since the 1980s, the healthcare system has been attempting to identify ways to share data beyond the individual provider's interaction. From the earliest attempts in the 1980s around Community Health Information Networks (CHINs) to Local Health Information Infrastructures (LHIIs) and then Regional Health Information Infrastructures (RHIIIs) to Regional Health Information Organizations

(RHIOs) to the nationwide efforts today including the Nationwide Health Information Network (NHIN and NHIN 2), all efforts have had limited success. With the earliest attempts in the 1980s, technology was often the limiting factor. In the early 2000s, the LHII and RHII were able to show that technology had now caught up, but hurdles around data and transport standards limited sharing of information. In addition, governance issues in managing the relationships among the organizations involved also caused many initiatives to falter or fail. When organizations, such as RHIOs, were created to address the governance issues, additional challenges arose. These included issues around data ownership and reuse, where information resides, tracking access, and even concerns over medical liability of having access to information and not acting on it in a timely manner. All of these issues limited organizations' willingness to participate fully in health information exchanges.

In 2004, through an executive order, the Bush administration created the Office of the National Coordinator for Health Information Technology and appointed Dr. David Brailer as its first coordinator. Emphasis was placed on leveraging health information technology to achieve many qualitative goals including the following:

- Ensuring secure and protected patient health information
- Improving health care quality
- Reducing health care costs
- Informing medical decisions at the time/place of care
- Improving coordination of care and information among hospitals, labs, physicians
- Improving public health activities and facilitates early identification/rapid response to public health emergencies
- Facilitating health and clinical research
- Promoting early detection, prevention, and management of chronic diseases
- Promoting a more effective marketplace
- Improving efforts to reduce health disparities

The majority of these goals require both the coordination of data and the effective, practical application of analytics. In addition, these goals are now focusing attention in deriving clinical and business value from information collected and applied across the clinical spectrum.

The clinical spectrum is the interaction between the research and practical sides of healthcare. As noted in the diagram below, problems requiring some form of treatment are identified in the healthcare setting. They are then evaluated from both a clinical and business perspective in the best way—and even if—to research a treatment, usually a drug or similar medication. Once the research process starts, it has limited interaction with the larger healthcare environment until its relatively late stages. Once approved, the treatment is applied in practice by physicians in the care of their patients. Often, there is limited feedback on the effectiveness or adverse events of a treatment once it is officially in the market and actively used. This often results in unforeseen negative effects of a treatment being addressed too slowly or if at all. In addition, new opportunities for existing treatments or variations of them can go unresearched because of an inability to identify the opportunity due to distributed or unavailable data.

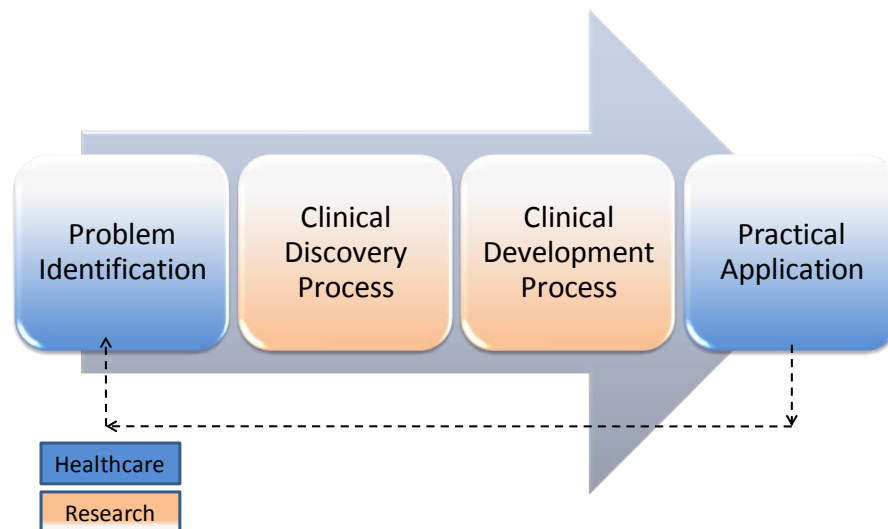


Figure 5 Treatment Development Process

NO MAN IS AN ISLAND – BUT WHAT ABOUT HIS DATA?

Returning to the current reactive healthcare model, the care management process usually starts only after a critical threshold is reached. This threshold is often the presentation of symptoms or, unfortunately, an event requiring immediate attention. There are a series of phases that an individual goes through during the care management process. Below is a list of the phases starting with indications of risk for a condition through diagnosis, treatment, and follow-up.



Figure 6 The Care Management Process

In each of these phases, there has been some attempt at analytics based on available data to help either in the identification, treatment, or management of the patient. These have ranged from fairly popular analytics such as health risk assessments to condition identification tools to identification of gaps in care to personalized care plans. The limiting factor in most of these analytics is the data. Most data used in current healthcare analytics suffers from one or more of the following challenges:

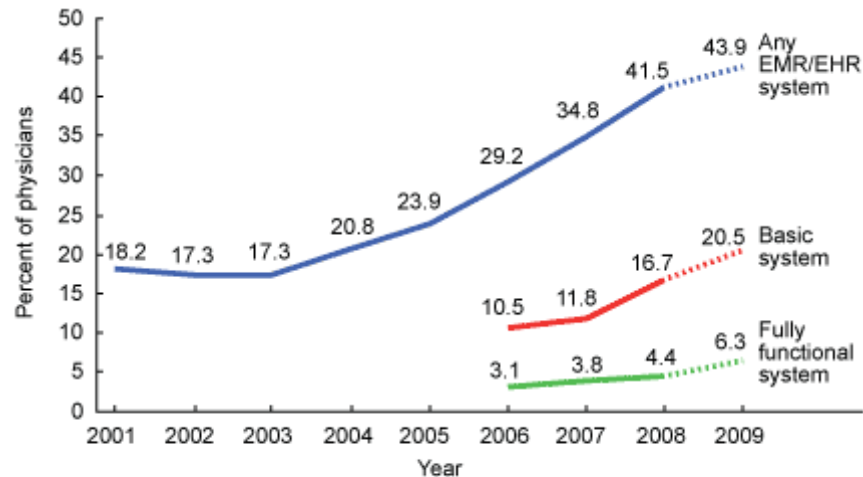
- Incomplete
 - Most data is captured in distributed systems focused on collecting specific information around a specific issue or functional need. These health information systems usually reside in different care settings. A complete record is unavailable because these systems are usually not accessible for aggregation of data and, when they are, often use different mechanisms for identifying an individual making it virtually impossible to be sure that all records for an individual can be linked together.
- Inaccurate
 - Many pieces of health information are captured on paper or verbally and then later transcribed into an electronic system. The more steps between the collection of information and the entry into a system, the greater the likelihood of error introduction.
- Self-reported
 - Much of the clinical or health observation information collected that isn't based on claims or prescription data is usually self-reported. This information, which might be collected in electronic health records or health risk assessments, is suspect since it is based on the interpretation of the individual and might be inaccurately recalled when being collected. Also, information is usually interpreted when delivered from a physician to an individual; so rather than specific data, the individual might only know the results such as they have high blood pressure or hypertension but not know the exact values or severity.
- Invalid

- Information collected and evaluated for these processes might be invalid due either to the nature in which the information was collected or how the information was reported. For example, often a code might be entered on a claims form (where much of today's electronic health information resides) in a way that either intentionally or accidentally might misrepresent the actual procedure that the individual received. Also, much of this information must be interpreted based on codes, which can lead to an incorrect recreation of the health care event.
- Claims-based
 - The majority of health care information today that is in an analysis-ready or shareable state is based on claims submitted to either health plans or to the government. This information focuses on the information necessary to process financial payment and usually does not include the clinical values needed to accurately assess the problem in terms of its clinical metrics (such as lab values). In essence, while it is possible to assess with some level of confidence what occurred and why, it is nearly impossible to assess how a treatment impacted the patient.
- Stale
 - Since the majority of information captured electronically does come from financial claims information, this information is usually several months old by the time it is accessible for analytical evaluation. This is due to both the time delay in submitting and adjudicating the claims. While great progress has been made in reducing the time to process claims and by the very nature that a claim can only be filed after services have been rendered means that this information can never be as current as other information that is collected during the care management process.
- Non-standardized
 - While there are a variety of standards for various types of information such as procedures, diagnoses and medications, there isn't widespread adoption of a single standard for the majority of information collected. While there are organizations such as the Healthcare Information Technology Standards Panel (HITSP) focused on harmonizing standards, the majority of health information is collected today in a way that makes it difficult, if not impossible, to reconcile similar information from different data sources.
- Distributed
 - Healthcare is becoming more and more distributed in its application. For a single healthcare event, an individual might interact with literally dozens of different healthcare providers and systems. It would not be unusual for a person to start out visiting a health care clinic and then being sent to a hospital where he or she would have seen at least one physician in each setting. He or she could then be referred to a specialist, have a battery of tests performed at a lab and need to visit a local pharmacy for medications. A severe event might require a stay in a skilled nursing facility or require periodic visits from a visiting nurse or social worker. Information collected as part of all of these interactions are usually stored in separate health information systems if stored at all. These separate systems rarely are linked; and if they are, they rarely will use a common approach to storing their information or referring to the patient. Aggregating information for an individual requires that all parties engaged in the healthcare process share information in some sort of standard way and have a common approach to identifying patients and providers. Below is an example of some of the myriad locations where health data for a single event might be located.



Figure 7 Health Data Locations for a Healthcare Event

- Non-electronic
 - Despite a number of efforts to encourage health information technology adoption by providers, the majority of patients' information is still captured and stored in a non-electronic, usually paper-based, system. While there are many estimates of health information technology adoption, the CDC recently put physician adoption of EMR/EHR systems that deliver basic EMR/EHR functionality at approximately 17% in 2008 and 21% in 2009. The reasons for this low level of adoption range from the cost to adopt new technology to the cost and burden to transfer old records into an electronic format to the lack of additional value most providers believe they will gain with a new system. To date, there are very few tools that can add enough value to a physician's daily experience with their patients that compensates for the cost and hassle to transform their current method of operating. In addition, there is a potential downside that many physicians see regarding transferring their systems into an electronic system and potentially sharing this information with others. This downside is a loss of what is often a competitive advantage for physicians in today's healthcare market—the fact that they do have your health information and others don't. A healthcare consumer is far more likely to continue to visit their current physician because their physician has their health information. Visiting a new physician would mean that your health history would need to be rebuilt, usually by the consumer conveying it through a "clipboard" medical history questionnaire. Diagnoses, medications, and previous procedures would only be known if the consumer was able to remember and convey all of this information. Should your entire health record be available online to any physician, healthcare consumers would be far more likely to shop around for a new provider.



NOTES: Any EMR/EHR is a medical or health record system that is either all or partially electronic (excluding systems solely for billing). The 2009 data are preliminary estimates (as shown on dashed lines), based only on the mail survey. Estimates of basic and fully functional systems prior to 2006 could not be computed because some items were not collected in the survey. Starting in 2007, the skip pattern after the all or partial EMR/EHR systems question was removed. Includes nonfederal, office-based physicians. Excludes radiologists, anesthesiologists, and pathologists.
SOURCE: CDC/NCHS, National Ambulatory Medical Care Survey.

Figure 8 Percentage of office-based physicians using electronic medical records/electronic health records (EMRs/EHRs): United States, 2001-2008 and preliminary 2009²

HEALTH CARE ANALYTICS – THE FIRST CHAPTER

Even with all the problems of getting healthcare data, a wave of analytic tools appeared within healthcare, usually within specific phases of the healthcare spectrum and usually driven by cost-savings rather than improved quality of care. Health payors such as commercial health plans or the government looked to using analytics to reduce the costs of their most costly users. These costly healthcare consumers were usually individuals with one or more chronic conditions such as heart failure, diabetes, asthma, or COPD. The costliest individuals were those that were either unaware of their condition or uneducated in how to manage their condition. Payors quickly realized that they could save a lot of money if they could get these individuals under a care management plan that helped them get their conditions in control and be proactive in identifying warning signs and seeking care when appropriate rather than waiting for an event and going to the hospital. Out of this initiative, analytics sprouted in the following key areas:

- Healthcare assessments
 - Often using either claims information, manually entered patient questionnaires, or both, healthcare assessments were used to help understand a patient's health history and their risks and to look for any indications that an individual might have or be at risk for a chronic condition. These assessments were suspect based on the data that they were using and their reliance on self-reported information. Often these assessments were used as a way to identify a sub-population from a group that would have further assessments.
- Risk stratification
 - Risk stratification tools were used to help organizations identify which individuals were most at risk for having a chronic condition or some sort of predictable healthcare event. Risk stratification tools were also limited by the availability of data and the fact that most of the information evaluated was self-reported.
- Condition identification
 - As a result of a health assessment or through the evaluation of claims information, analytics were often used to identify people with conditions who were either already diagnosed but unmanaged or who weren't diagnosed but had indications of a condition. Condition identification helped payors focus on what part of their population would benefit from chronic care management.

² From the CDC website: http://www.cdc.gov/nchs/data/hestat/emr_ehr/emr_ehr.htm

- Condition severity
 - In an effort to apply the appropriate resources to the appropriate people, analytics was used to help assess both the severity of the condition and the degree to which it was under control. Organizations with care management programs would often provide a “high touch” approach to their sickest and most out of control patients while offering a “low touch” approach (such as mailings or self-directed online programs) to their lower severity or better managed patients. Interestingly enough, it was later demonstrated that while almost all patients in a chronic care management program received a health benefit by being enrolled, there was very little difference in the benefit between those with high- versus low-touch approaches.
- Gaps of care identification
 - In an effort to understand how to help physicians and care managers help those with chronic conditions better control their condition, evaluations of claims information and information collected via Web-based or phone-based interviews were used to identify care items—such as assessments, medication regimens, or procedures – that might help the individual either improve their health status or be proactive in caring for themselves.
- Care Management predictive modeling
 - Although rudimentary compared to analytics available today, the original care management providers often looked at helping health plans and employers evaluate the potential pool of care management participants to identify which individuals would likely receive the biggest benefit and, by extension, save the payor or employer the most amount of money. In addition, these predictive modeling tools would attempt to assess the right number of individuals to manage against the number of care managers needed to manage them. The resulting ROI would help guide the sponsors of these care management programs with both the population to manage as well as the type of intervention (based on cost) to apply to each participant.
- Care plan creation
 - One of the last places analytics has been applied in healthcare was in aiding the development and execution of care plans. While care plans were sometimes created by a health payor-sponsored care manager or by the patient’s physician, analytics was used to help determine what care plan items would help and in what order they should be prioritized. Some of the more sophisticated algorithms employed additional dimensions in their evaluations such as which care plan items would likely have the most impact on the patient’s health and which care plan items would most likely be adopted or adhered to by the patient. These more sophisticated models attempted to make the care plan more personalized based on the individual patient’s situation and preferences.

The graphic below highlights a sample of analytics offered at various places along the healthcare spectrum.

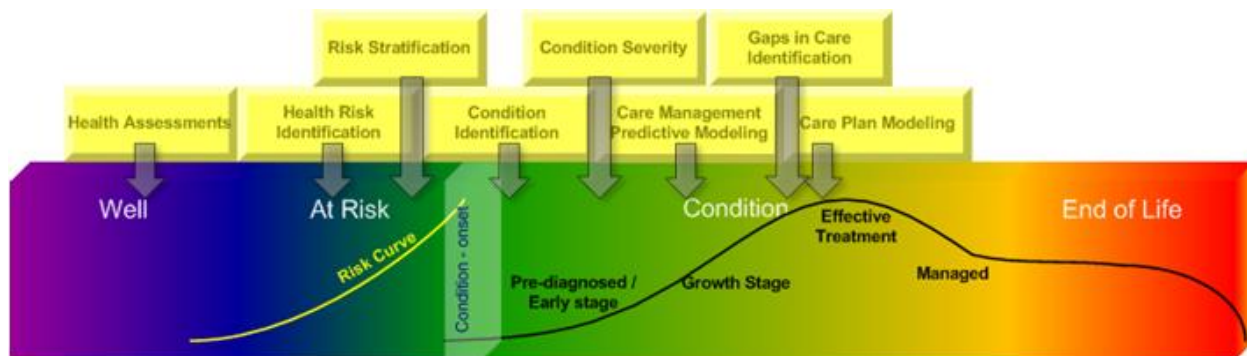


Figure 9 Analytics Applied in Care Management

There has also been a fair amount of analytics applied outside of the care management process. Two examples include physician profiling (the analysis of how well a physician is adhering to certain quality metrics) and claims analysis (the analysis of evaluating claims for incorrect coding or healthcare fraud).

IT'S ALL ABOUT YOU - FINALLY

There's an old riddle that goes "There's something that the person who makes it doesn't need, the person who buys it doesn't want, and the person who uses it doesn't know it. What is it?" Your health information is very similar. Your doctor evaluates you, the lab runs tests and provides the data back to the doctor, the physician consults with the health plan to determine which treatments are acceptable, a prescription is given to the pharmacy who dispenses the medication, and the payor decides how much to pay for each of the billable items. The information that's collected and shared among these parties is almost never shared directly with you; each party involved believes to some extent that they own the data that they generated, regardless that it's about you. Even though HIPAA grants each individual the right to request their information, it is rarely given readily or in an easy to understand format.

Finally, there is an initiative underway to encourage the adoption of health information into longitudinal health records around you. Before now, the health information technology initiatives underway were mainly focused on sharing information among providers, not with the patient or healthcare consumer. Amazingly, most of the RHIOs and Health Information Exchanges created had little to no consumer representation. Sometimes this was due to the fact that it was challenging to properly represent an amorphous concept as "the patient.". More often this was due to the fact that these organizations were more focused on reducing the administrative burden in sharing components of health information rather than creating a comprehensive record about you or improving the quality of care delivered.

The push from the federal government to expedite the adoption of health information technology that delivers meaningful use (including electronic health records that support certain standards) will help health information to be shared more easily and eventually leveraged for research and practical purposes. In the Life Sciences industry, standards such as those proposed by organizations like CDISC are helping to make clinical trial information more available for sharing and aggregating. Combining information from the clinical research with observational health data will finally enable the creation of cross-industry research databases that provide the capabilities to analyze information longitudinally around three critical dimensions in health care—the condition, the treatment, and the individual.

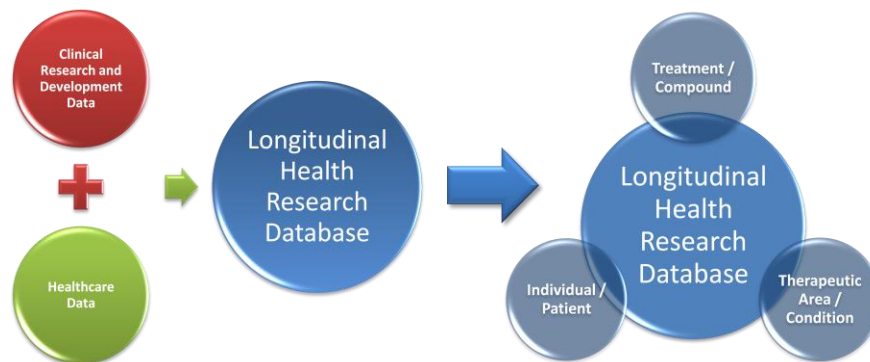


Figure 10 Combining Health Information from the Health and Life Sciences Industry Enables Research by Condition, Treatment, and Individual

By the way, for those of you who were wondering, the answer to the riddle is a coffin. I'll let you decide if that's ironic.

THE DEMAND FOR INTELLIGENT HEALTHCARE AND THE ADVANCED ANALYTICS TO SUPPORT IT

For those of you following closely the healthcare debate currently raging through the various parts of the U.S. government, you will have noticed that about mid-2009, the emphasis changed from healthcare reform to health insurance reform. For those of us who have been struggling to apply analytics throughout the larger healthcare environment, this was somewhat disheartening. However, recently with funding from ARRA and initiatives from ONCHIT, there is an underground movement carrying the mantle of analytics-driven healthcare. Buzzwords like Personalized Medicine, Comparative Effectiveness and Meaningful Use have begun to pervade the latest round of sponsored initiatives. These types of initiatives can only be undertaken with some form of a comprehensive longitudinal research database and advanced analytics. Personalized Medicine is dependent on both the detail data of an individual including in many cases their genetic blueprint and the effectiveness of various treatments both in research and healthcare settings. This information then needs to be accessible at the point of care to the provider to allow an informed decision about the various treatment options and the likely potential outcomes. While nothing will ever replace the personal care and expertise of your physician, the practical application of personalized medicine will enable a more efficient delivery of the latest research to the point of care based not just on the recommendations for a generalized patient, but an individual matching the characteristics and situation of you, the patient.

The sister initiative to Personalized Medicine is Comparative Effectiveness. For years, health plans and other payors have been making deals with pharmaceutical companies to get discounts on drugs and medications in exchange for putting them on a formulary. Doctors routinely use this formulary for prescribing medications to their patients, often choosing the one that the health plan has on formulary rather than a more costly version that might produce better long-term benefits. Comparative Effectiveness initiatives will help change the debate from cost to value. In essence, different drugs and treatments might be more beneficial than others in either subsets of populations or when evaluated against a long-term payoff. A longitudinal view of individuals and their health outcomes is needed to understand which populations benefit from which treatments, both in terms of health outcomes and cost. Advanced analytics is needed to evaluate the specific comparisons between treatment options and also to evaluate for an individual which comparisons are relevant for his or her experience.

ANALYTICS IN THE CONVERGING HEALTH AND LIFE SCIENCES INDUSTRIES

With the potential aggregation of a cross-industry longitudinal health information database, comes enormous opportunity for new or enhanced analytic solutions that can truly be the catalysts that change both the research and practice of healthcare. In addition, assuming health information collected in the two industries can be shared in near-real-time, the following is a partial list of some new or enhanced advanced analytics and how they benefit the individual and the healthcare system.

- Personalized Medicine (including genomics and phenotypic evaluation)
 - As mentioned above, personalized medicine involves the analytics of evaluating the specific treatment options available for the individual based on his or her specific characteristics and the effectiveness of treatments relevant to his or her specific situation.
- Comparative Effectiveness
 - An evaluation of all individuals using specific treatments from a value perspective both in terms of improvement in health quality as well as long-term cost. Effectiveness is also compared at various degrees of granularity to evaluate whether specific treatments are more effective based on certain patient characteristics.
- Clinical Trial Recruitment
 - Analytics that allow for both real-time recruitment of patients into clinical trials based on specific characteristics of an ideal participant as well as analytics that evaluate an available patient population to proactively identify potential clinical trial participants based on an ideal participant profile.
- Biosurveillance (Bioterrorism monitoring)
 - Real-time analysis of health information captured from all healthcare data sources looking for indications of a potential bioterrorism threat.
- Signal detection for patient safety (adverse event monitoring)
 - Analytics looking at longitudinal health information as well as mining social media sites looking for signals of unreported or underreported serious adverse events.
- Clinical Trial Simulation
 - Analytics evaluating a comprehensive longitudinal health information database for simulating clinical trials using cohorts selected from the aggregated population based on specific participant characteristics.
- Signal detection for new indications
 - Evaluation of all users of a particular treatment, drug, or compound looking for any unexpected new indications as a result of the use of the treatment alone or in combination with other treatments or drugs.
- Retrospective Clinical Trials
 - Evaluation of users of marketed drugs to compare safety and efficacy results with those collected during the drug development process.
- Physician Profiling

- Evaluation of providers along a number of potential dimensions, such as adherence to guidelines, qualitative improvements for various conditions, breadth of treatment options, and success rates for procedures.
- Condition Profiling
 - Evaluation and analysis of a condition based on the characteristics of the individuals who have the condition or the characteristics, if applicable, of the target condition itself (such as the clinical analysis of specific cancerous tumors).
- Patient Profiling
 - Evaluation and analysis of patients including their risks, conditions, care plan strategy, and treatment adoption and adherence likelihood.
- Proactive Healthcare
 - The evaluation of individuals and their health and lifestyle behavior to predict potential negative health events and address with interventions as early as possible. The goal of Proactive Healthcare being to keep the individual in a “well” state as long as possible while quickly reacting to any health events with early and effective interventions with the goal of reducing the time it takes to recover from the event.

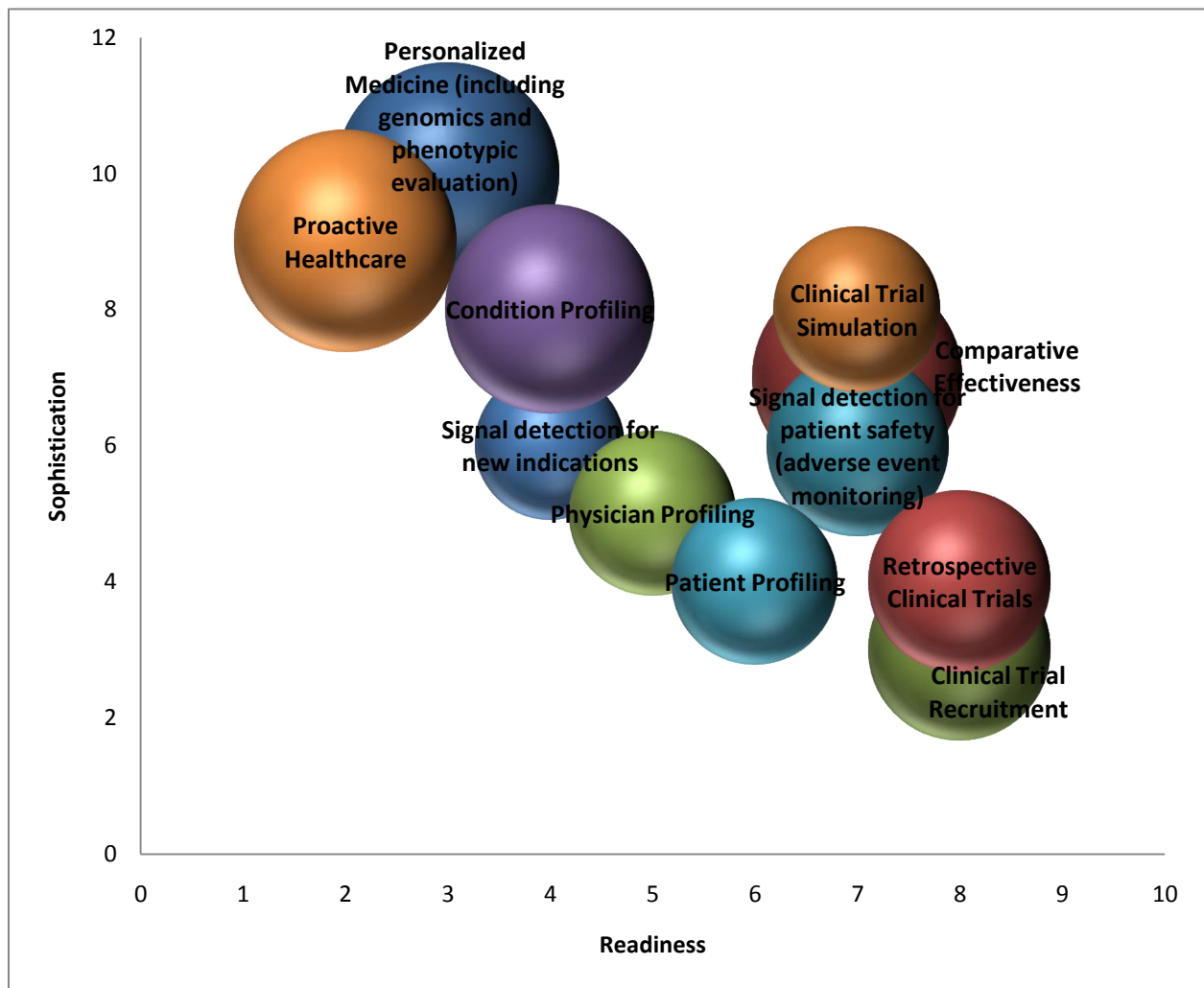


Figure 11 Potential Advanced Analytic Solutions with a Converged Health and Life Sciences Health Information Data Store (size of bubble indicates potential impact)

INTERMEDIATE ANALYTICS

The analytic solutions listed above are hampered from achieving their full potential until some immediate challenges are addressed. These challenges can be addressed with analytic solutions that are either available today or can be developed using existing analytic capabilities. In general, these challenges fall into the following five categories:

- Data Transformation and Standardization
 - With data still being stored in siloed, proprietary health information systems or, even worse, on paper, effort is needed to create data adapters that can map information from these stores into a flexible and extensible data standard. The ability to handle both structured and unstructured data as well as transform unstructured data into searchable content must be handled in order to establish a comprehensive data store.
- Data Aggregation and De-identification
 - Once data has been transformed into a common standard, a way to aggregate the data into either a centralized store or a series of smaller, but federated data stores is needed. The information must be able to be viewed and analyzed as an aggregate, population level data set, as a de-identified data set for individual comparisons or, if needed for authorized activities, able to be linked back to its identifying characteristics.
- Data Confidence and Validity
 - Analytics to help define the confidence level of the data's validity and integrity. As various analytics might only want to use information that can be defended in these regards, information will need to keep track of how it was captured, how it was entered, and if it can be corroborated with other information or by multiple sources. Without some confidence mechanism, both clinical research and healthcare providers will be suspicious of the information on which they're basing their evaluations and recommendations.
- Master Data Management / Master Entity Index
 - Given the distributed nature of healthcare and the fact that even within a single organization an individual patient, provider or medication might be stored with a multitude of different unique identifiers, it will be up to an analytical process during data aggregation to provide a reconciliation mechanism to link or associate all records related to a specific entity such as a patient, physician, or compound.
- Episode Grouping
 - The last of the major hurdles will be a way to evaluate all the events related to a health event and combine them into a logical group for future analytic evaluation. The episode grouping will have to take into account potential duplicate records from multiple systems as well as ascertaining how to split apart multiple events that happen to occur simultaneously.

THE FUTURE IS NOW

While many of the topics presented in this paper seem idealistic and futuristic, the fact is that many of these initiatives are currently underway. ARRA is spurring many comparative effectiveness research initiatives by providing funding to the following agencies who are sponsoring comparative effectiveness projects:

- \$300 million for the Agency for Healthcare Research and Quality
- \$400 million for the National Institutes of Health
- \$400 million for the Office of the Secretary of Health and Human Services

The National Cancer Institute is currently in the midst of developing a National Cancer Outcomes Database to house all research and health information related to cancer research or treatment. The goal is to help both researchers and healthcare providers be able to understand the disease better, gain a better understanding of the cancer patients, and look for more effective ways to develop and execute clinical trials. Dr. John E. Niederhuber, Director, NCI, has explicitly stated that it is his goal that every cancer patient should be identified and incorporated into an appropriate clinical trial. NCI is in an ideal position to provide a holistic data store of research and health information due to their ability to both fund research and collaborate with a number of academic and cancer clinics.

The effort to leverage analytics on a comprehensive clinical research and healthcare database isn't limited to the US. In the United Kingdom, the National Health Service (NHS) is continuing to evaluate funding the Research Capability Programme (RCP). As stated in its vision statement,

The Vision of the NHS Research Capability Programme (a collaboration between the National Institute for Health Research and NHS Connecting for Health) is:

To enable better health outcomes for the public and patients achieved at best value for the taxpayer; and
 To support the ambition to make the UK the preferred place to carry out medical research, by building a nationwide health data and information platform that will enable health research to achieve its maximum potential.

It will achieve this by:

Providing a common information infrastructure that will be adaptable, and develop in response to the research process and the needs of all stakeholders (including patients);

Providing a customer-focused set of services that both ensure information is treated and handled in a safe and secure way to protect the rights and confidentiality of patients and other identifiable individuals under the law, and also meet researchers' needs efficiently. The services will enable high quality research directed towards improving health outcomes: effective treatments; patient safety; and quality of life; and

Providing access to a comprehensive range of technical resources and data sets, under strict protocols of information governance, that will aid the research community to access the data needed to conduct research of the highest methodological standard.³

Even SAS is joining in the effort. While SAS is actively participating in the NCI initiative and is offering assistance to the Secretary of Health and Human Services for their Comparative Effectiveness Research initiative, SAS is also looking at evaluating its own unique circumstance by looking at applying analytics to health information collected for over 30 years from the SAS Health Care Center. While still in its initial exploration stage, SAS understands that there are very few opportunities to have a longitudinal data set of health records for individuals spanning as long as three decades. Analysis of the outcomes, costs, and absenteeism of the SAS population in comparison to other populations are being considered in an effort led by Gale Adcock, Director, Corporate Health Services for SAS. The goal of the initiative is to help understand what benefits have been realized by the SAS Health Care Center and how these can be shared in other health care settings.

CONCLUSION

With health reform focusing on moving from reactive to proactive healthcare while at the same time making sure that those already addressing a health issue get the appropriate treatment based on their own personal characteristics, the demand and desire for analytics-driven healthcare is growing. The technology limitations of the past are no longer an issue. With the government focused on aligning incentives for adopting and leveraging health information technology for delivering meaningful care, the environment is now ready for the integration of data from both the Life Sciences and Healthcare industries. With a comprehensive data set, advanced analytics can now be applied to better understand individuals, conditions, and treatments both in terms of specific comparisons and long-term perspectives. The availability of advanced analytics for modeling, simulation, and forecasting allows both the Life Sciences and Healthcare industries to focus their energies in researching and delivering more effective strategies for preventing, managing, and curing individuals' health issues. Individual consumers, with a comprehensive electronic health record at their disposal, will be able to leverage a variety of analytic tools to better understand their current and potential health future in a way that allows them to be truly informed and active participants in the healthcare system rather than passive bystanders with minimal access to health professionals.

With the practical application of advanced analytics in the converged health and life sciences space, the purpose of healthcare can finally change from "Getting Better" to "Staying Well."

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Matt Gross
 Enterprise: SAS
 Address: SAS Campus Drive, S2034

³ Research Capability Programme – Enabling Phase NPFIT-RCP-INTR-0001 in PD00 Background, Vision, Overview and Glossary 13August 2008 vPV Publication Version

City, State ZIP: Cary, NC 27513
Work Phone: (919) 531-2959
E-mail: matt.gross@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

Other brand and product names are trademarks of their respective companies.