Predictive Analytics: The Future of Value-Based Healthcare

The triple goals of greater access, better economic efficiency, and better outcomes are increasingly served by predictive analytics. This white paper explains some important use cases that are being solved using predictive analytics. Then it outlines some of the key challenges that occur within the core business processes that take place when creating and operating a predictive analytics program. Finally, it outlines reference data and functional architectures that you can customize for your own organization. Upon reading this paper, you should be able to get started crafting a predictive analytics program and choosing partners who can ensure your success.

**Predictive Analytics Presents Important Use Cases Driving Costs Down and Quality Up**

Healthcare presents the perfect storm for predictive analytics. The digitalization of the clinical record offers vast quantities of information. Innovations in data management allow for unstructured clinical notes to be organized into compelling structures. Healthcare terminology engines abound to make this extraction of facts from clinical and administrative repositories simple. Increasing numbers of technologists understand the nuances of how to manage and organize data and workflow so that creating and integrating predictions into healthcare workflow is becoming increasingly simple. Finally, we find that major healthcare providers and technology vendors often partner together to create predictive solutions and bring them to market. Let’s look at some of the interesting use cases that demonstrate these trends.

- **Cleveland Clinic creates a simple risk adjustment score to evaluate provider quality**. In assessing the quality of care delivered, creating an apples-to-apples comparison within a population is impossible; no two health conditions are exactly alike. To solve this problem, creating a risk adjustment score for individuals who suffer similar but unique situations allows for statistical models to be predictive. Using sparsely annotated procedure codes, the Cleveland Clinic was able to compare factors not related to a patient’s physiology across populations. The power of this study is in its ability to compare employers, health plans, and institutions while accounting for the fact that the populations they represent are composed of unique individuals. The innovation uses simple data available even from administrative records.

- **Providence Health demonstrates value of multidisciplinary teams collaborating to manage high-risk patients**. Providence Health Plan has created a simple financial model to determine which patients will be enrolled in a care management program. Providence Health Plan’s disease management programs include congestive heart failure, coronary artery disease, diabetes
management, chronic obstructive pulmonary disease (COPD), and asthma. On a per-disease basis, the institution is able to assess risk based on the amount of money spent in an ambulatory or hospital setting. While the algorithm is crude, this use case shows once again how choosing simple data to operate on sometimes can yield important results. As care teams collaborate with data scientists, working toward simple and elegant solutions is often sufficient when more complex data mining is not possible.

- **Dartmouth-Hitchcock predicts readmission risk.** In a coordinated set of activities, including the creation of readmission predictive models culled from Epic Clarity’s data warehouse, Dartmouth remained one of the 22% of American hospitals that was able to avoid any readmission penalties levied by the Centers for Medicare and Medicaid Services (CMS). For congestive heart failure, heart attack, pneumonia, knee or hip replacements, and lung ailments such as chronic bronchitis, Dartmouth-Hitchcock is in the top 2% of hospitals in terms of its excellence in avoiding unnecessary readmissions.

> “It really is a whole system we’ve put in place to ensure that our patients are healing once they have left the hospital. This has been an incremental effort to continue our excellence of care.” —Darlene Saler, administrative director for patient flow and care transitions

- **Google uses unlikely data sources to predict in semi-real time where the flu and dengue are.** Google has a team of data scientists and collaborates with teams of epidemiologists from all over the world to parse the search streams in order to detect in real-time where the flu and dengue are. In the case of the flu, there is great surveillance data that the team uses to validate its findings. In the case of dengue, Google is often the best source of epidemiological data that exists in many developing countries. Based on the search stream, Google predicts where the flu actually is, and months later the team gets to validate its predictions from more traditional surveillance methods. In this use case, we see a fundamental epidemiological breakthrough taking place. Suddenly we don’t need to wait months to see disease trends. In the years ahead, this type of out-of-the-box thinking will determine to some extent which healthcare systems dominate in the new world where predictions are embedded into clinical and administrative software.

**CHALLENGES MUST BE OVERCOME AT ALL POINTS IN THE PROCESS**

Implementing an effective predictive analytics program is not simple, but with good change management processes in place, a body of best practices can be put into play to ensure success. The cultural challenges, though often soft, are the hardest to cope with. Will the clinical staff feel that they are losing the ability to practice the art of medicine? Can the predictive insight be integrated into the medical workflow effectively and without the clinical staff having to do complex activities to take advantage of it? Can the privacy concerns of the individuals whose data is being analyzed be truly and legally protectediii? It is only in considering these sorts of cultural issues that a program will
become successful. In addition to that, we document some of the challenges that must be overcome for each major step in the workflow.

Figure 1: The core workflow for predictive analytics

- **Establish and Refine Goals:** The most important activity in any predictive analytics program is to establish the goals to be accomplished. Driving a predictive analytics program from real and important business cases allows a team to discover what data speaks to these goals as factors to consider. Then predictive models can be built based on these hypotheses and institutional capabilities.

- **Examine and Curate Data:** Once goals are established, look to the data that can likely be predictive and then curate it. Understand its quality and how it must be cleaned and structured in order to be actionable. Then, if your team is working with good tools, finding the nuances within the big picture and building a proper model from the software used should be almost automatic.

- **Close Data Gaps:** It will probably be true that, once your team understands the data that is easily available, there will be gaps in the data in terms of its optimal ability to be predictive. For example, in creating standards of care via predictive mining of data, it is often necessary to understand what each node on a care path costs. But most health systems are not equipped to
do this. However, doing this work via curation is not overwhelming and adds great value. So often a more clearly predictive model can be derived by licensing supplemental data, creating metadata, or mining hard-to-digest nuggets.

- **Test and Refine Hypotheses:** Finding the right model and algorithm is almost always iterative. Therefore, do not seek perfection on the first pass; seek sufficient predictive power to take action. Realize better care, more access, or more efficiency with this initial data, rather than trying to create the perfect predictive model before taking action. But also in parallel, reduce the noise from bad signals, add more metadata to explain the meaning of the raw data, and ensure that the machine learning algorithms your solution uses are effectively tuning themselves so that the model and results are more accurate.

- **Make Findings Actionable:** Once you have a set of predictions, pilot making the findings actionable. Ensure that the predictions and the interventions your organization does in relation to the predictions make a great deal of sense in a pilot before making them actionable. Ideally do peer-reviewed research, ensuring the rigor of methods used to test the effectiveness of the interventions of your organization.

- **Operationalize Workflows:** Once the intervention path is established as effective, it is urgent to incorporate that into the clinical workflow. Most modern electronic medical records (EMR) systems allow for call-outs to remote processes — for example, a call-out to a clinical recommendation engine that could incorporate the predictions in order to suggest sets of clinical orders that mitigate the risks raised by the healthcare predictions. It is important that your institution does not go to the expense of rolling out these integrations before you have established their effectiveness. Also it is urgent that as you integrate the predictions into the clinical workflow, you consider the end-to-end experience of system users; it must not be more difficult to action the predictions than to neglect them, or your institution will find them often neglected.
**Predictive Analytics for Providers: A Reference Architecture**

Below we show two dimensions of a reference architecture for predictive analytics in healthcare. First we speak about the types of data that are most likely to be used in doing healthcare predictions. Then we go on to a simple version of a functional or capability architecture to demonstrate what business capabilities are necessary to do this work.

**Data Architecture**

The data architecture is divided into four knowledge domains. For the most part, each source of data on the right of the chart is derived from one of the knowledge domains. However, this is not clear-cut. For example, much of the research data could be from any of the other domains, and much of the social data might come from any of a number of repositories. However, the key point here is that in order to predict something in one of the knowledge domains, it is often necessary to understand facts from another domain. For example, if a clinical recommendation engine were aware of the reimbursement policies of the patient’s insurance plan, including such things as formulary rules, then the engine would not recommend procedures or tests that could not be carried out because the patient could not pay for them.

![Knowledge Domain Diagram](image)

*Figure 2: Clinical, administrative, social, and research data are often combined to create predictive insights.*

Below we discuss all four of the knowledge domains separately.
- **Clinical Domain:** The clinical domain yields information from many sources. The distinction of the information from this domain is that it is all protected health information (PHI) and thus subject to the rules of HIPAA if it relates to any individual. Thus it must be secured and handled with extreme care. Sources for this type of data are labs, images and reports and findings around the image studies, EMRs, any monitoring equipment that is hooked into the system, and information from other providers in a circle of care.

- **Administrative Domain:** All of the financial information around claims benefits and other matters are the administrative data. This data is often embedded with many clinical facts, such as diagnosis and procedure code. Often this is the best data an institution can find to build predictive models from and is therefore of great value. When this information is coupled with financial facts, such as the patient’s financial responsibility as well as total costs for a billable action, it can be overwhelmingly valuable.

- **Social Domain:** This is a large domain and includes all the information about the socioeconomic environment that an individual is from. So, for example, in predicting who might be defrauding an institution, it is often a relevant social fact that a provider inhabits a building owned by a known criminal. The state of a person’s relationship to the welfare system or being homeless are other social facts of great importance when predicting risk. Data mined from social media sites such as Facebook will play an increasingly important role in establishing powerful predictive models.

- **Research Domain:** The research domain is in many ways the most complex but fertile. Institutions like Mount Sinai’s Icahn Institute for Genomics and MultiScale Biology are spending hundreds of millions of dollars to create new methods of research where complex interactions of genes with the environment can predict new drug solutions for complex disease states. One advantage of the research domain is that funding is often available to do complex data management tasks. One disadvantage is that the mining of research data is often a daunting and frustrating task.

**Functional Architecture**

Below is a simple functional architecture showing the business capabilities necessary to create healthcare predictions. Many of these capabilities are generic to good data management practices, but a number of them are specialized to the ability to do the conative tasks of finding complex patterns via machine learning and other artificial intelligence capabilities. Not depicted are the skills necessary to determine what the program goals are and how to measure success, but a well-organized program will not neglect these capabilities.
Create Predictions

<table>
<thead>
<tr>
<th>Manage Data Sources</th>
<th>Manage Meta Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Administrative Data</td>
<td>• Data Structures</td>
</tr>
<tr>
<td>• Clinical Data</td>
<td>• Rules</td>
</tr>
<tr>
<td>• Social Data</td>
<td>• Performance indicators</td>
</tr>
<tr>
<td>• Research Data</td>
<td>• Algorithms</td>
</tr>
<tr>
<td>• Device Signals</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Segment Population</th>
<th>Manipulate Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Assign Individuals Risk Scores</td>
<td>• Extract Data from Unstructured Sources</td>
</tr>
<tr>
<td>• Assign Individuals Behavior Change Propensity Scores</td>
<td>• Define Key Performance Indicators</td>
</tr>
<tr>
<td>• Assign Individual to Cohort, or predictive group</td>
<td>• Find Patterns for Predictive Models</td>
</tr>
<tr>
<td></td>
<td>• Eliminate Noise from data</td>
</tr>
<tr>
<td></td>
<td>• Operate on Date, and numerical data mathematically</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predict Health Facts</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Extract predictions from Pattern algorithms</td>
</tr>
<tr>
<td>• Ensure that traceable reason(s) can be presented to users</td>
</tr>
<tr>
<td>• Publish Predictions, or Exposed them in real time</td>
</tr>
<tr>
<td>• Integrate Predictions with core workflow tools</td>
</tr>
</tbody>
</table>

**Figure 3: The functional architecture of predictive analytics for healthcare**

Below we list the major functional areas that are required to do predictive analytics for healthcare.

- **Manage Data Sources**: To manage a data source, it is important to understand its quality and what the operational pathway is to get this data. Sometimes the data will come from a warehouse, a staging platform, or a data mart that is owned by a particular application. Sometimes the data will come from systems. One of the most important things to think about as you design your production systems is change management; that is, for your upstream data sources, how can you anticipate change and not be encumbered by coming changes? To get to the heart of this data, it is often necessary to have a medical terminology engine that can parse and structure data in a specialize form of extract, transform, and load (ETL) that is knowledgeable with the medical domains.

- **Manage Metadata**: The data about the data is the metadata. For example, the ICD-10 codes have particular meanings, and knowing the diagnosis codes alone is often not sufficient to understand risks. Rather, in that most risk for expensive or dangerous health conditions comes from co-morbidity interactions, understanding the relationship between diseases is almost
always necessary to create good predictive models. For institutions that are not capable of understanding these relationships, various specifications, such as the CMS DRG codes, can be of great help.

- **Segment Population:** At the core of almost all healthcare predictive analytics is the need to create population segments. Even when the goal is to evaluate a provider, understanding the populations of patients that a provider sees is a core determinant. Thus we list the creation of population segments as a core capability unique to healthcare analytics. Two elements of this are important: Using the physiological state of a patient as well as it is known from the data sources is primary. Also important is the need to understand how to segment a population based on its ability to respond to interventions. If a person is at great risk, but there is nothing that anybody can do about that risk, then it is probably not worthwhile to expend a great deal of resources trying to get that person to change a behavior, as the person will not do so and the result will be frustration from both the caregiver and the patient.

- **Manipulate Data:** A core requirement is that the data can be mined, transformed, matched, and managed via most ETL functions and, usually with healthcare data, with aggressive capabilities to do natural language processing wherever unstructured data is of importance. It is increasingly easy to get tools to do this, and as terminology engines become more sophisticated, these capabilities are embedded within them.

- **Predict Health Facts:** This is the culminating capability. Core to this capability are two nonobvious functions: It is essential that any solution around predicting what a system should do about human health be able to express how it came to that prediction. Sometimes this is as simple as presenting a clear algorithm, like “Who spent more than 15K in ambulatory care last year?” and sometimes it can be extremely involved when explaining complex networks of dependence in order to come to a predictive conclusion. The other important capability is to integrate the predictions with the other systems of workflow and control so that they can be actioned in the everyday operational world of the caregivers or, if they are being exposed to consumers, with an interface that consumers find convenient and simple.

**Recommendations**

I. Ensure that you establish the business goals of any predictive work that the institution undertakes. Solve important problems, and be willing to spend time and money to do so. This is something that can succeed only if the data scientists, the technologists, the bioinformatics staff, and the clinical staff work together in order to get results. Marshaling this many stakeholders only can be accomplished if there are institutional goals that folks can rally around.

II. Start with simple algorithms and data feeds. Make progress off good enough. Don’t seek perfection of data quality or predictive output.
III. In order to ensure that a rigor of method exists, consider getting funding for scientific study to be peer reviewed. If your institution does this, then rolling it out operationally will be an easier task because the research will speak more loudly than vendors’ boasts or sponsors’ pride.

IV. Once it is time to make a predictive set of facts operational, carefully consider the end-to-end workflow of the clinic and do sufficient integration with core systems so that the workforce does not feel burdened in adapting to these important changes.
About the Author
Skip Snow, is an independent industry analyst for the software that drives healthcare. His website can be found at www.hipaabox.com.

Skip is a seasoned technologist, analyst, and entrepreneur with a history of success pioneering application architecture for some of the largest corporations in the world. Most recently he served as the senior healthcare analyst for technology at Forrester Research. His first IT job was in the mid-'90s where while working for Columbia University's medical school, he built one of the first interactive patient care systems using the Web's then primitive text-driven interface, allowing nurses to both enter data into a patient’s electronic chart and look up care protocols from the same application. Later, after years doing development, and architecture in the banking sector, Skip served as vice president of technology architecture at Kaiser Permanente, the largest nonprofit health maintenance organization (HMO) in the US. In this role, Skip was charged with developing the information, security, and infrastructure for Kaiser’s enterprise.

Skip was a pioneer of online banking and service-oriented architecture (SOA) at Citigroup, ending his tenure there as the chief SOA architect. He developed a software-as-a-service (SaaS) back-up service for medical data as the CEO of HIPAA Box, and for the past several years he has consulted around the topic of IT strategy and enterprise architecture for clients such as Johnson and Johnson and IMS Health.

Skip is the former chair of the Web Service Interoperability Organization's Standards Group (WS-I) and a member of the World Wide Web Consortium's (W3) Web Service Policy working group. He is no stranger to the startup world, and is regularly consulted by venture firms for his insight regarding enterprise technology. A native of New York, Skip now lives and works in Los Angeles, CA.

About Rapid Insight
Rapid Insight Inc. is a leading provider of predictive analytics software and solutions that provides organizations with the ability to make data-driven decisions. Focusing on speed, efficiency, and usability, Rapid Insight products enable users of any skill level to quickly turn their raw data into actionable information. The company’s analytic software platform simplifies the extraction, analysis, reporting, and modeling of data for clients ranging from small businesses to Fortune 500 companies.

For more information, visit: www.rapidinsightinc.com.
http://my.clevelandclinic.org/services/anesthesiology/outcomes-research/risk-quantification-index. Creating a risk-adjusted model is the key to evaluating provider quality. In their study “Development and validation of a risk quantification index for 30-day postoperative mortality and morbidity in noncardiac surgical patients,” Dalton JE et al studied more than 600,000 patient records to create and validate a risk quantification index used in predicting morbidity and other outcomes of cardiac patients.

See this report on the predictive models and processes of intervention planning based on simple financial facts.

In this presentation from Rock Health, a venture capital fund specializing in healthcare software out of Silicon Valley, these cultural barriers are discussed and addressed.

for a list of the DRG codes. Deeper analysis will map these codes to risk.