ADVANCED ANALYTICS FOR EFFICIENT HEALTHCARE

Data-Driven Scheduling to Reduce No-Shows

dataiku

Summary

Introduction: One Dataset a Day Won't Keep the Doctor Away

The healthcare industry is currently suffering from a host of issues. Knowledge sharing between hospitals, determination of patient adherence to medications, and the efficient management of surgical procedures are just three topics in a long list of areas that need improvement. All of these issues have the same thing in common: the healthcare industry has a data problem.

The fact is that there is an abundance of raw data and no one really knows what to do with it. From patient records to heart-rate monitors, hospitals produce reams of raw data that, after an initial reference, is usually forgotten. The good news is that all of this data can be used to solve a multitude of common, day-to-day problems using predictive analytics.

In this ebook we'll highlight a specific issue—no-show appointments—and show how predictive analytics can be used to discover real-world solutions to a multi-billion dollar problem.

From patient records to heart-rate monitors, hospitals produce reams of raw data that, after an initial reference, is usually forgotten

The No-Show Problem

The unfortunate reality is that no-shows have become extremely common — one study reported that the no-show rate in U.S. primary care practice can vary from as little as 5% to as much as 55% 1.Appointment cancellation rates are also a systemic problem in addiction treatment centers where 29% - 42% of patients fail to begin treatment 2 and 15% - 50% of patients do not even return for a second visit 3.

Dealing effectively with patient no-shows has been a challenge in the healthcare industry, especially now that reimbursement is more closely tied to performance measures surrounding physical appointments.

The long-term effect of this phenomenon is lowered reimbursement for providers and, more importantly, the health welfare impact on adherence, quality, and clinical outcome measures on patients. For patients, spotty appearances with healthcare providers results in less coordinated care, particularly in cases of chronic diseases and preventive encounters. Patients suffering from chronic conditions may require very regimented treatment plans — missing even one treatment may have debilitating consequences.

Missing preventive care treatments leads to longer and more expensive care as potential issues become real health problems. No-shows also have a direct financial effect on healthcare providers as expected revenue targets fall short, labor hours are wasted, and inefficiencies are created.

The challenge has always been, "What do we do with all this data? How do we add meaning to it?"

Dealing effectively with patient no-shows has been a challenge in the healthcare industry, especially now that reimbursement is more closely tied to performance measures surrounding physical appointments. **Many providers are simply**

overwhelmed with the problem and resort to traditional stopgap policies, such as reminding patients the day before their appointments. The effect is marginal and, ultimately, is short-sighted because it does not directly address the problem itself.

Executive Summary

This ebook aims at providing healthcare professionals with a clear view of how efficiency gains could be realized at little cost via the integration of data analytics. We will start by having a look at what is wrong with the current implementation of data analytics in the healthcare ecosystem and how it applies to the no-show issue. We will then offer an alternative approach to addressing no-show appointments that makes use of predictive analytics. Lastly, we will discuss how this method could be applied to the healthcare industry.

CHAPTER 1

Data Fragmentation and Limited Skills Deteriorate the Data Analysis Process

The healthcare industry is no stranger to data — they have an abundance of it, from clinical measures and demographic information to lab results and staffing data . Figuring out what to do with all of the data is a challenge that lies at the heart of medicine in the 21st century. According to a recent survey released by the National Association of ACOs (NAACOS), 51% of Medicare Shared Savings Program (MSSP) ranked data-related issues, such as access, inconsistency, and deployment, as the biggest roadblocks in their accountable care journey 4. Core obstacles of data management & usage include:

- *translating data into actionable information for providers,*
- *acquiring the required skill-sets needed to analyze the data,*
- *and finding solutions that can report on the business aspect of clinical data.*

> *Apart from the structural dysfunctions, healthcare suffers at the IT level: technologies are out-of-date compared to other high-tech sectors, institutions use proprietary platforms that are incompatible with other systems, and the IT skill-sets of employees are highly disparate.*

The "data issue" is exacerbated by the nature of the U.S. healthcare ecosystem: it is a highly fragmented industry across multiple sectors. Healthcare is rarely coordinated, incentives are misaligned, and variation is ubiquitous.

Apart from the structural dysfunctions, healthcare suffers at the IT level*: technologies are out-of-date compared to other high-tech sectors, institutions use proprietary platforms that are incompatible with other systems, and the IT skill-sets of employees are highly disparate.*

Multiplicity of Data Sources Makes Collection & Use Difficult

Too Much Data

As mentioned, there is an abundance of data in the healthcare industry and all of it flows from multiple sources: 5

- Prescription, diagnostic (lab, vitals measurements), and demographic sources;
- Social media / Web-based and machine-to-machine data (e.g., remote devices);
- Transactional data (e.g., claims and billing activities);
- Biometric data; and,
- Human-generated data (e.g., Electronic Health Records [EHR], physician notes).

According to the Institute for Health Technology Transformation, the amount of U.S. healthcare data reached 150 exabytes in 2012 and is estimated to double by 2022 6. One exabyte equals 1 billion gigabytes and, given that the human brain can only process around 7 variables at once, it's obvious that the sheer size of the data involved poses a significant challenge.

Collecting the Data

A large percentage of human-generated and biometric data is transcribed into Electronic Health Record (EHR) systems; many of these platforms were invented in the 1960s and, frequently, little has changed in the basic approach used to categorize incoming data. These one-size-fits-all systems are not well-suited to individual workflows and they also lack the personalization needed to truly understand the needs of the patient.

Familiar Data Sources

Many organizations use data sources that are comfortable, familiar, and accessible. Over time the usage of these data sources become increasingly entrenched in healthcare environments, to the point where other sources of data are not even considered. The problem with this approach is that it only provides a partial picture and does not provide access to the value that big data analytics can offer.

Data Diversity Hinders Data Integration

Lack of Standards and Use of Multiple Standalone Silos of Data

In U.S. hospitals, the documentation of incoming data is mandated via the use of EHR solutions. Ideally, an organization would use a common data entry interface for all departments, from the emergency room to the finance division. Such a framework would enable an analytics solution to access multiple data point originations for comparison and analysis, effectively providing a holistic view at the operations & management levels.
The reality, however, is much different: currently, 72% of healthcare organizations use more than 10 electronic interfaces to collect data.

This level of disparity between data sources is a product of environments that use individual silos of data: the accounting department collects data their way, patient biometrics are collected a different way, and so on. Consequently, there is no real standardization of data across an organization.

In addition, single-function EHR systems do not have the capability to aggregate, transform, or create actionable analytics. **In fact, intelligence is largely delegated to retrospective reporting which is insufficient for forward-looking healthcare data analytics initiatives.**

No Real-Time Data Integration

There may be some healthcare organizations with advanced data collection capabilities, but there are few that possess advanced data integration at the intra and inter-organization level. Meaning, there are no mechanisms in place to support the sharing of data between healthcare institutions. There is an anecdotal story about one hospital that was unable to share data with another hospital located just across the street — data had to be printed and manually entered into the other hospital's EHR.

At the U.S. government level, there is much concern over a plan to share the EHRs of 10 million military service members from hundreds of hospitals and clinics across multiple public & private agencies — a monumental task estimated to cost at least $11 billion over a decade [7].

> *The focus of decision-makers—in terms of the use of data—has traditionally been applied to identifying volume & cost trends within fiscal reporting periods rather than the actual use of real-time data at the operations level.*

Data integration issues are also present within healthcare institutions. For example, **most internal HIT (Hospital Information Technology) systems do not offer real-time data APIs**; typically, this data is processed overnight and available in the data warehouse on the following day. The avoidance of offering real-time data analysis is indicative of the overall approach of the healthcare ecosystem to data.

The reality is that organizations are rarely data-driven — there is little internal incentive to evangelize real-time data analysis. The reasons for this stem from the nature of healthcare: administration-level decisions must take into account a host of contractual, regulatory, and political decisions before being implemented.

In addition, the focus of decision-makers—in terms of the use of data—has traditionally been applied to identifying volume & cost trends within fiscal reporting periods rather than the actual use of real-time data at the operations level.

Limited Human Skills Inhibit Effective Data Analysis

Few Data Skills

In terms of data analysis at the human resources level, there is a severe disconnect between the skills required and the skills currently available. Healthcare organizations require data professionals with a range of skills that are not solely technical. Today's data scientists need to expand their skill-sets to include soft skills such as communication, collaboration, creativity, and leadership. It is not enough for a data scientist to know how to design and build analytical models — they must be able to work with their peers to add meaning to the data and then successfully convey that information to healthcare professionals who do not have an IT background. **A July 2014 survey of healthcare leaders stated that 60% of them were unsure of whether their organizations had the in-house expertise necessary; in many cases, system development skill-sets were outsourced 8.**

In addition to the type of skill-sets needed, there is a shortage of talent. The McKinsey Global Institute estimates that there will be a 10,000+ analytic talent shortage through 2020; the end-result of this shortage means that 50% - 60% of data scientist positions could go unfilled 9.

> *It is not enough for a data scientist to know how to design and build analytical models — they must be able to work with their peers.*

Reactive vs ProActive: A Shift in Industry Culture

As organizations grow and mature, there is an increasing reliance on established methodologies — even if those methodologies are inefficient. Long-standing procedures are given merit simply due to their age; in these situations, "this is the way it's always been done" is a common refrain.

The shift to an analytics-based culture is a significant one because it requires the abandonment of long-standing reactive procedures and the adoption of a proactive data-driven approach. In this environment, organizations use a single source of truth to guide choices, they stop making "gut decisions," and avoid data shopping (i.e., finding data that supports conclusions that have already been made).

CHAPTER 2

Data Management Challenges Illustrated by the No-Show Issue

The ubiquity of no-shows has put a spotlight on a set of broader data management issues in the healthcare industry. The inability of healthcare organizations to deal with the no-show issue has had a profound effect on patient health, their experiences with healthcare providers, and on the financial bottom line. The problem is a difficult one to solve due largely to industry practices that are both archaic and ineffective.

The Financial and Human Cost behind No-Shows

A Costly Reality: Upward of $150b per Year

When a patient is unable to attend an appointment, there are multiple repercussions that affect much more than the healthcare provider's bank account. The U.S. healthcare system loses more than $150 billion per year in no-shows alone; these **costs stem largely from all of the associated issues that come into play when a patient is unable to attend an appointment** 10.

For example, missing an appointment means that the overhead related to that appointment is not reimbursed — items such as staffing costs, insurance, and utilities remain on the books. In addition, a significant number of appointments are made on a referral-basis — cancellations made at the primary care level means that those referrals are never made, while cancellations at the specialist level means that more revenue is lost and the patient's health may suffer. Ultimately, no-shows have a significant impact on everyone, from physicians to patients, as physician costs increase in order to bridge the financial gap caused by missed appointments.

Example

- A doctor is supposed to see 15 patients every day;
- 10% no-show rate = 1,5 missed appointments daily = 8 no-shows per week
- The doctor organizes appointments into 30-minute sessions at a cost of $150/session.

Because of the 10% no-show rate, he loses $1,200 per week. This no-show rate costs the practice around $62,400 per year.

Beyond the Financial Cost: Deterioration of the Patient Experience

An unintended consequence of no-shows are negative patient experiences (e.g., long waits or abbreviated visits). This is due to attempts by healthcare organizations to solve the no-show problem using ill-conceived methods.

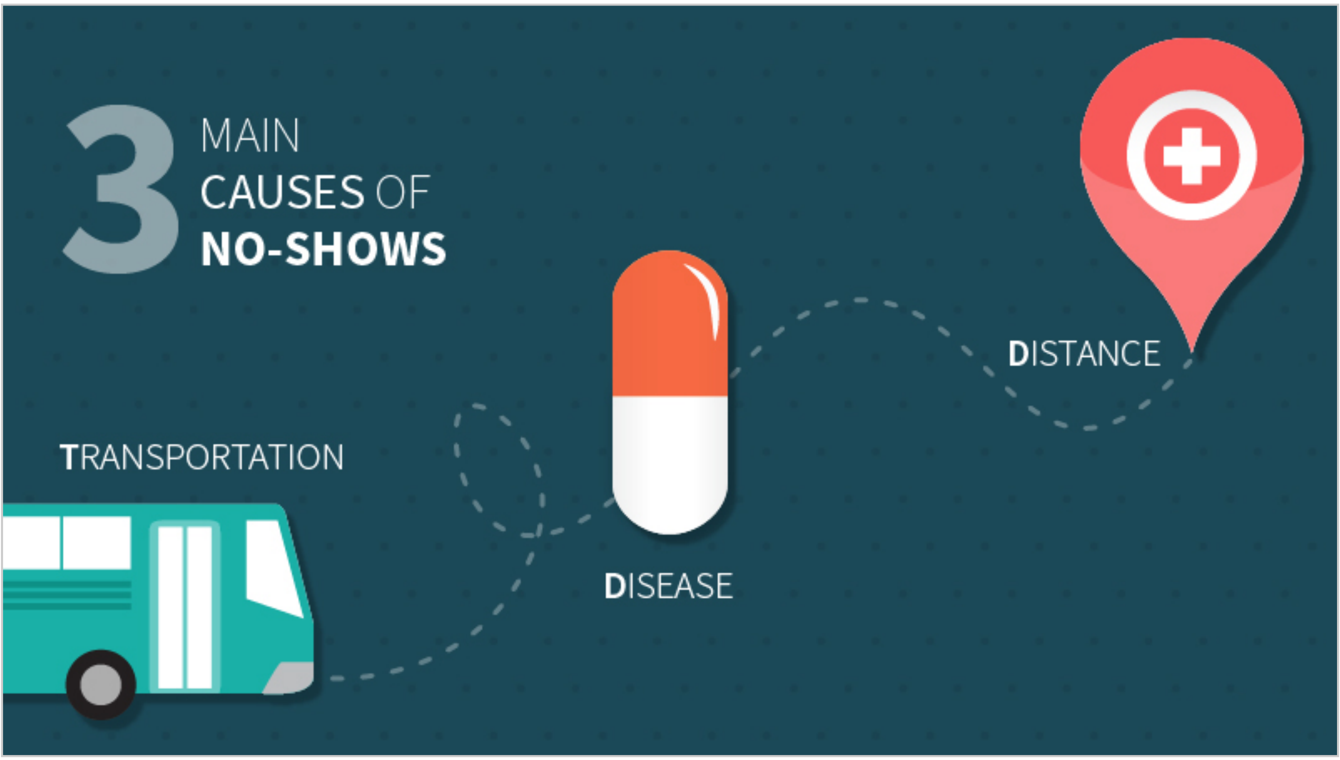
For example, some health centers implement financial penalties for missed appointments. Another technique is to double-book patients; this results in short appointments (e.g., 15 minutes instead of 30 minutes) that do not give patients an opportunity to properly address their health concerns.

The people affected the most by no-shows are really the patients themselves. Preventive healthcare is responsible for discovering a wide array of potentially life-threatening diseases, but is frequently eschewed when patients either procrastinate or simply decide to cancel their appointments. For example, diabetic patients with weakened immune systems can use wound clinics to treat minor cuts… this does not cost much. If they decide to cancel their appointment, though, a small issue may turn into much larger (and more expensive) problem.

The 3 Main Barriers to Solving the No-Show Issue

Difficulty with the Incorporation of Data Sources

Why do patients cancel appointments? There are many factors that are responsible for no-shows and it's not always about something else "coming up."
Some of these factors include distance (i.e., geographically remote), transportation (e.g., lack of public transportation options), and scheduling (i.e., too early, too late, etc.). Sometimes appointment destinations are selected based on patient data that may be outdated. The type and severity of the disease being treated is also a no-show factor.



In terms of predictive analytics, all of these factors represent data points that could be used to decrease the likelihood of no-shows or, at least, bet-

ter anticipate them. The reality, however, is that HIT systems typically lack the capability to take advantage of this data and are unable to access data stored across different silos.

Lack of Access to Patient Data

Frequently the scheduler, or the person responsible for scheduling doctor-patient appointments, does not have access to the global patient dataset, which means that the scheduler may be working with either a lack of data or even incorrect data. This may be due to a lack of technical knowledge on the part of the scheduler or it may be due to the underlying system being used (non-collaborative). **The end-result is that the scheduler cannot see the "big picture,"** such as the patient's transportation needs, exact geographic location, and time availability. This lack of knowledge means that the scheduler is effectively working blind and is unable to create appointments that reflect the patient's needs.

Lack of No-Show History

Profiling no-show patients is a difficult task because there is not too much relevant data readily available. This is particularly true for patients with a health plan who have no history of seeing a doctor within their network; **predictions are quite difficult because there is no historical data to analyze.** The progressive incorporation of new data, along with external data sources, may provide the clues needed to determine which patients will likely not appear, but data systems that use such real-time data are not widespread.

4 Quick Fixes... That Don't Work

Double-Booking

Double-booking is when a healthcare practice sets appointments for more people than they can viably handle with the expectation that a percentage of them will not appear, effectively creating a fully-booked day with normal doctor-patient visit times. The high number of unknown variables, however, results in a very negative patient experience.

If there are no no-shows, then all patients only have access to their provider for 50% of the expected time (e.g., 15 minutes instead of 30 minutes). **As the day progresses, the staff and physicians become frustrated, often wishing that someone would not appear for their scheduled visit.** The patients themselves also become frustrated as they are rushed through their appointment

First-Come, First-Served

The first-come, first-served approach is a common method in 3rd world countries in order to alleviate the need for scheduling maintenance. The idea is that care is given to whoever shows up first; a ranking system is kept by the front desk personnel based on the patient's order of appearance.

The primary issue with this approach is that **healthcare providers frequently have little incentive to perform their job punctually**; after all, they know that their office will be full of eager patients every morning. This exacerbates the core problem, which is the long wait — patients are kept waiting for hours until it is their turn (and until the physician appears).

Financial Penalties

In an effort to negatively incentivize a patient's on-time appearance, some healthcare providers implement a financial penalty to patients who do not appear (or appear late). This obviously has **negative repercussions on patients from all walks of life, as those without money will be unable to pay the penalty**.

Contractual Appointment Reminder Service

Appointment reminder services are typically call centers hired on a contractual basis whose job is to call patients 24 hours before their scheduled appointment and provide a reminder. The average no-show rate is 12%. This method decreases the no-show rate to 10%. **The effect is minimal, the costs can be high, and the experience is impersonal.**

As a whole, all of the above methods are reactive by nature (trying to mitigate the problem) instead of being proactive (directly solving the core problem).

Interlude: Guidelines to Conquer the No-Show Issue

A Painful Issue

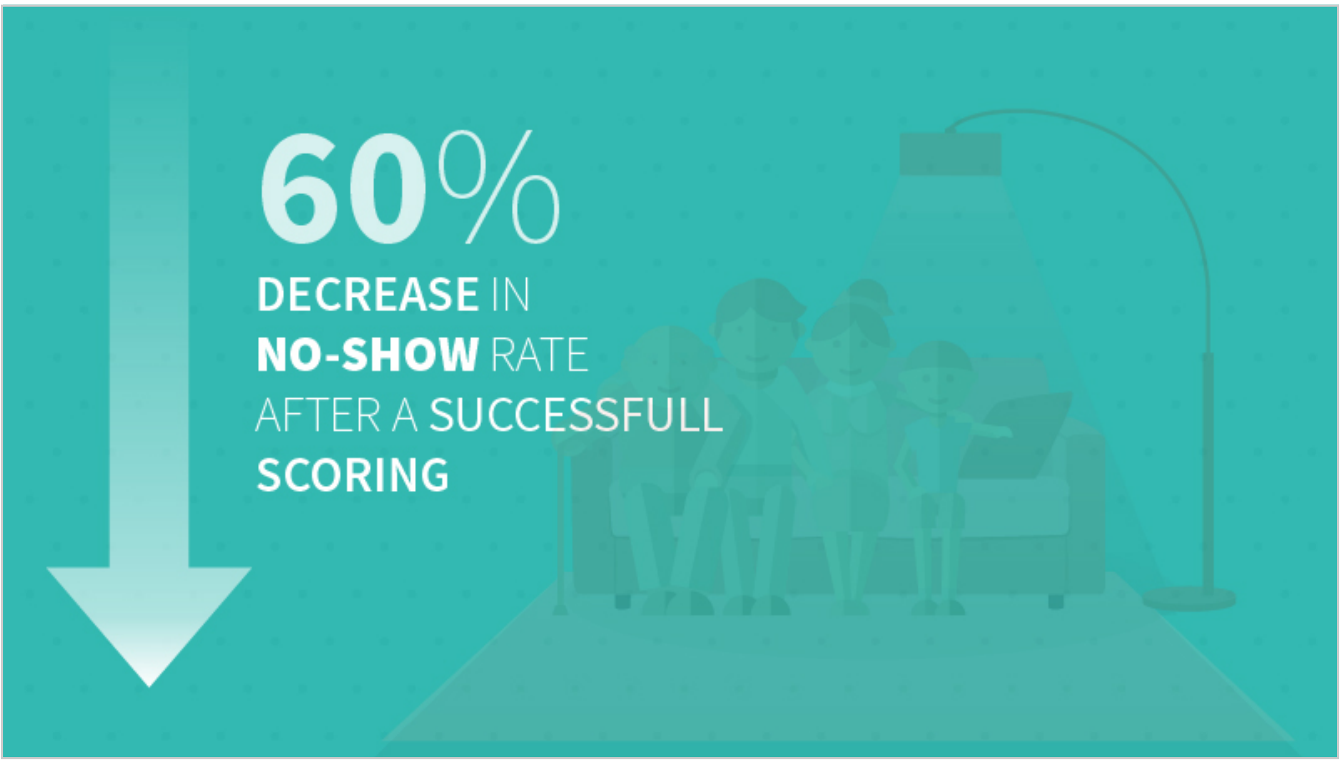
It's been a typical, and frustrating, day. It's 5pm and 12% of today's 300 scheduled patients did not show up for their appointments. This means that 36 people did not appear and your staff worked to 88% of their capabilities. **At this rate, you've been losing about $5,400 per day ($1.36 million annually) plus payroll & infrastructure costs.**

At the end of the day, you realize that you've been wasting money, frustrating your staff, and losing efficiency. Your healthcare payers are dissatisfied with your hospital's efficiency and the patients they are in charge of are not well cared for.

A Solution

Now, imagine if we could somehow reduce the 12% no-show rate by scoring the patient likelihood of a no-show. For example, a scoring mechanism could isolate the 5% of your patients that represent 40% of those most likely to not appear for their appointment. Instead of using a reminder service to call all patients, which costs both time and money, why not allow your scheduling staff to contact specific patients, remind them of their appointment and, if needed, arrange more flexibility. **This would reduce your no-show rate down to 7% and would save your hospital $550k per year — happy patients and a less frustrated staff.**

So, let's take this a step further. Instead of doing a one-time analysis of patient no-shows, imagine if you could use a predictive analytics methodology to determine no-shows in real-time.



Processing your Data

The process may start with a **computation of datasets to determine which times have the highest no-show rates.** This analysis may provide some surprising insights into exactly when your patients are not appearing. Secondly, given the local time slot data combined with global dimensional data, determine the reasons why patients are not appearing for specific time slots — possible culprits could be the weather, the geography, the disease, transportation options, and/or the patient. Defining these items would enable you to create & assign time-based points to each of your patients, depending on their distinctive features.

Deploying your Strategy

Armed with this knowledge, your schedulers could be advised, in real-time, on how likely your patients are to be a no-show based on time slot. **The scheduling process could suggest 3 specific time slots when patients' no-show sco-**

ring would be at its lowest. (i.e., when they are the most likely to appear). Combined with an overbooking strategy on specific time slots, this kind of proactive scheduling would enable your scheduling staff to suggest relevant time slots while also offering them flexibility. The end-result would be a no-show rate of only 4% and an annual savings of almost $1 million.

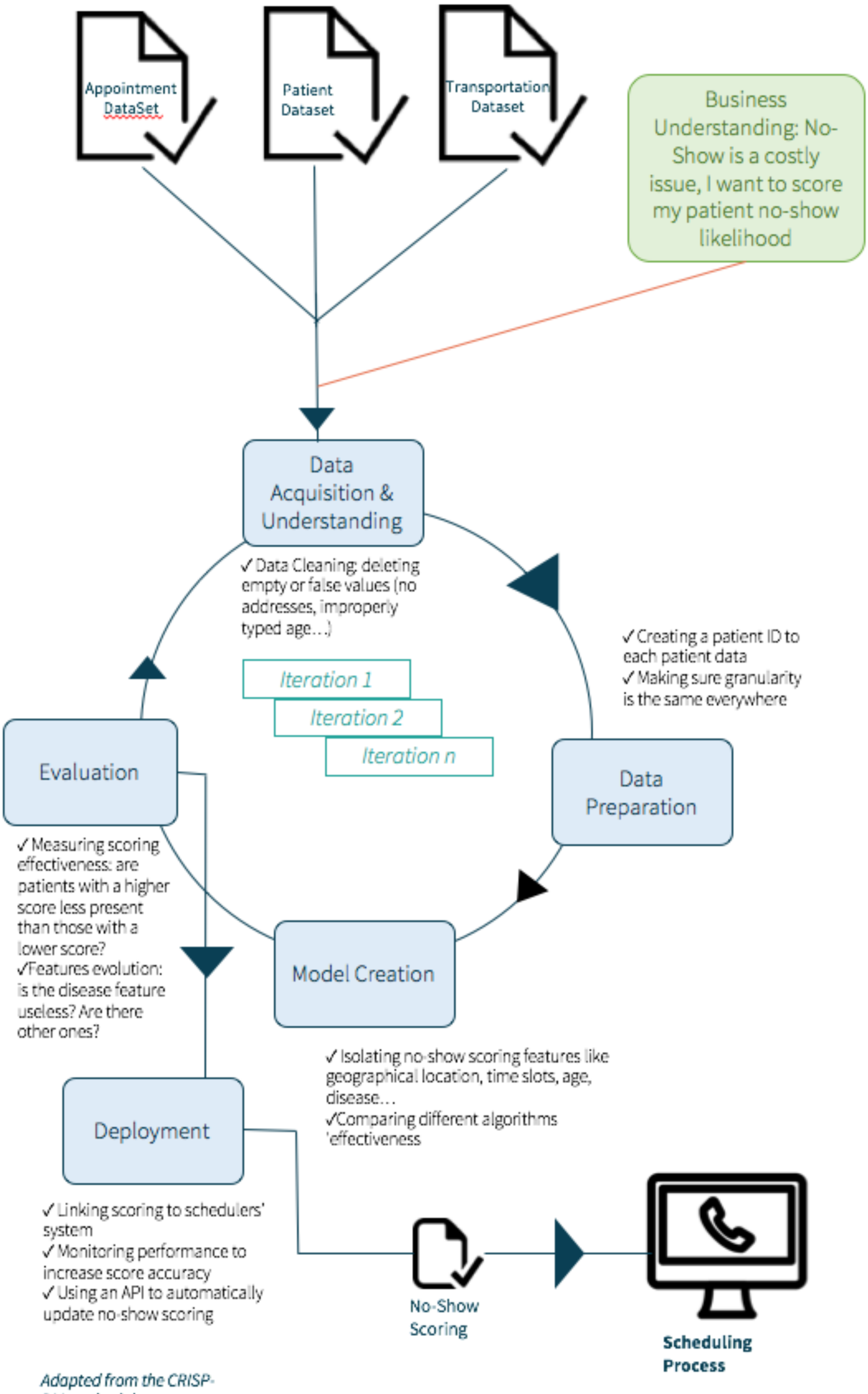
CHAPTER 3

Step by Step Methodology to Build your Scheduling Data Product

The process of developing a data analytics strategy to tackle the no-show problem requires a comprehensive methodology. ***The approach should start with defining a tangible goal and end with incorporating the output into business practices****; between those two end-points we have a complete agile-based process that includes data definition, gathering, cleaning, processing, and improving.*

Below is an agile roadmap that conveys how each process contributes to the eventual goal of deploying a predictive service to your scheduling system.

Six steps to a No-Show Scoring



Adapted from the CRISP-DM methodology

Defining the Perfect Framework

Designing the Perfect Frame

The fascinating aspect of data analytics and no-show problem solving is that all of this is possible. The data is there and the capability exists. Solving the no-show issue would promote efficiency across your organization while providing substantial cost savings for the long-term. Tackling this problem is not only a goal for healthcare providers, but also for healthcare payers who need to keep a grip on their population.
We will explore the possibilities below and discuss how data analytics can be used to negate the effect of no-shows and drive efficiency from the bottom-up.

First, though, we will discuss two critical aspects required when defining an effective project frame for healthcare analytics projects:

<u>Collaborative Framework</u>

As the saying goes, "**No man is an island**" — we are social beings who are most effective when cooperating and working with others. In terms of healthcare analytics, this means that Health Departments and IT Departments need to work hand-in-hand to effectively realize change. Likewise, an engaging analytics software solution needs to be collaborative and available to data experts as well as beginners.

<u>Agile Framework</u>

The Agile method is an iterative process whereby constant testing and incremental improvements lead to continuous improvements in the least amount of time. Agile frameworks are particularly well-suited to healthcare data analytics, because it enables your teams to constantly test models and prototypes in an efficient manner. **The Agile team should be inclusive and collaborative; members should be representative of both Health and IT Departments.** For example:

- Quality Director from the Medical Department;
- Data Scientists from the IT Department; and,
- Director of Primary Care from the Operations Department.

Order Out of Chaos: Collecting & Making Sense of Data

Define your Goal(s)

In order to keep costs within budget and to realize feasible results, **it is necessary to specifically define the project goal**. In this case, our goal is to score the likelihood of patient no-shows in real-time. The scoring would be used to identify high-risk patients and schedule the best time slots for them in order to decrease the likelihood of subsequent no-shows.

Collect Historical Data (Appointment Dataset)

In order to create an algorithm, the predictive analytics solution needs to work with data. If possible, **provide 3 months' worth of historical show/no-show data**; if not possible, you may need to collect this data for 3 months before beginning the predictive modeling process.

Gather Workable and Clean Datasets

Next, we need to determine the datasets that will be used to establish patient scoring. In other words, the factors that will determine whether or not a patient is likely to appear for a given time slot. Some possibilities include:

- Appointment Dataset: historical data of shows and no-shows;
- Patient Datasets: age, location, health problems, diseases, children, status...
- External Sources: social mapping of geographic area, transportation data, disease classification (i.e., effect of disease on the patient's lifestyle — for example, wheelchair-bound? mobility? capabilities? limitations?), bank holiday calendars, weather, and so on.

Some key questions to answer: how frequently are these datasets updated? Are they automated? Is accurate and up-to-date data available?

Combine and Clean your Sources

Combine all data sources, clean the data, delete empty/incorrect fields, and ensure that the same level of detail—in terms of granularity—is applied across all data points (e.g., weather data may be available daily while appointment sheets are created on a weekly basis). It is common for datasets to be available in different formats (xls, calendar files…), so one of the challenges of data collection will be shaping them all in a common processing-friendly format.

A Predictive Model to Test your Hypothesis

Highlight and Pinpoint Distinct Features

The process of building a predictive model involves a series of normalization and optimization steps designed to determine model accuracy. Some key steps in this process include feature normalization, testing & optimization of models, determination of model accuracy, and the specification of a user strategy. After the model is defined, the data scientist needs to overfit the model, evaluate, and ultimately validate it in order to isolate features.

The determination of accuracy is done by testing the underlying strategy in practice; for example, given patients who are likely to appear for a given time-slot, do they actually show up as expected? How accurate is the time-slot scoring for patients who do appear? If overbooking is implemented, is it being applied correctly? These questions all need to be addressed in order to determine the accuracy of the underlying analytical model — this involves comparing real-world results with the relevant predictions. This level of additional analysis will enable a data analytics solution to further refine the model's accuracy, if needed.

Of course if you are using an advanced software analytics solution, then many of the above steps would be automated. It would be able to clean datasets, isolate specific features, and automatically score the likelihood of patient no-shows.

These questions all need to be addressed in order to determine the accuracy of the underlying analytical model — this involves comparing real-world results with the relevant predictions.

Train Machine Learning Models on Test Datasets

If new features are added, then the models need to be re-trained. Additionally, data visualization needs to be done in order to determine if the features are relevant.

From Theory to Practice: Deploying your Data Product

Automate Data Preparation on Incoming Data

The hard work of data analytics is over. At this point, the outputs need to be incorporated into the scheduling process. **The first step of deployment is to automate the preparation of new incoming data** — this ensures that the solution continues to work effectively going forward.

Deploy Daily Results to Scheduling Team

After the predictive analytics solution creates the patient scores, there needs to be a mechanism in place to integrate the results into the scheduling system. **An API should be used to ensure that schedulers can easily access scoring.** The goal here is to score patients in real-time when schedulers are creating appointments. Ideally, multiple time slots should be presented to the scheduler so that there is room for scheduling flexibility. Each suggested time slot reflects the time when the patient is most likely to appear for the appointment.

The Reality Check

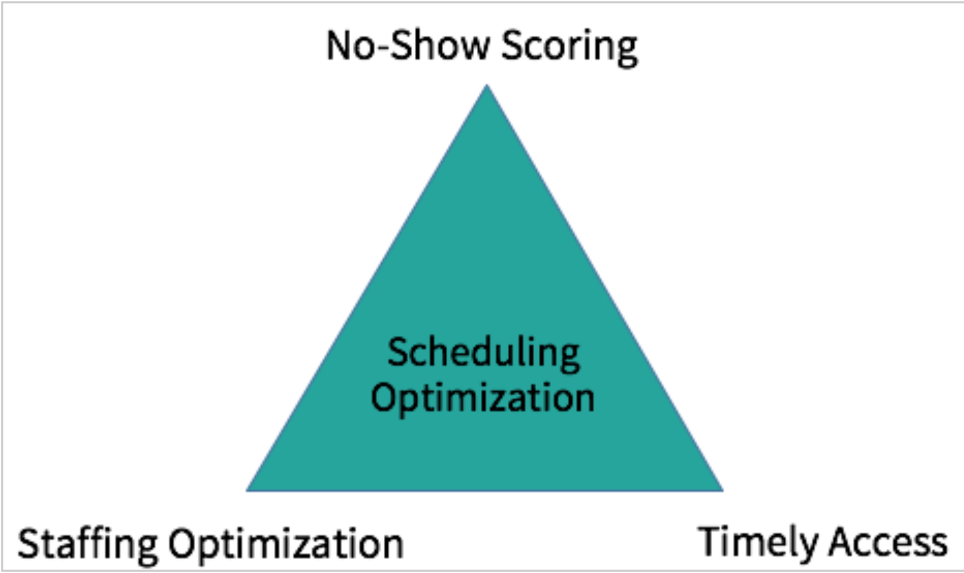
Once the model has determined which time slots have the lowest no-show likelihood, you are halfway there. Appointment scheduling lies at the intersection of efficiency and timely access to health services. **Timely access is important for realizing good medical outcomes and is also an important determinant of patient satisfaction.** For example, if three patients with near-identical no-show scores are scheduled for the same time slot, then the outcome may not be as expected: two of them will have to wait and the third will probably leave after half an hour.

Scheduling issues are magnified when considering staffing optimization. If multiple patients are scheduled for the same time slot, what will your staff do with their remaining work hours?

No-show issues have an impact on multiple areas across your organization, frequently in ways that are not expected. Establishing a scheduling optimization system means taking into account no-show scoring in combination with staffing optimization and timely access

One effective strategy is to develop an intelligent overbooking system based on specific time-slots in order to decrease the no-show rate as much as possible.

Your scheduling optimization model is all about being realistic: for obvious reasons, you can't schedule each of your patients on the same time slots. It's unlikely that your no-show rate will be equal to zero, so the best approach is to minimize it as much as possible. One effective strategy is to develop an intelligent overbooking system based on specific time-slots in order to decrease the no-show rate as much as possible. Such a system, powered by a machine learning approach, could be enriched with previous results so that the overbooking rate could be fine-tuned for specific time slots.


No-Show Scoring
Scheduling
Optimization
Staffing Optimization
Timely Access

Interlude - Predictive Analytics in Action

A major U.S. healthcare provider, responsible for 15+ hospitals and clinics, decided to deploy a predictive analytics solution in order to address the no-show issue. In their situation, the most important factors to determine patient appearance were time slots, location, and disease type. They discovered that there was a significant relationship between public transportation schedules and benign diseases. In order to address this, the provider always scheduled benign disease-related appointments in the middle of the day in order to sync with the availability of public transportation schedules.

They have now deployed real-time no-show scoring within their scheduling process. Three time slots are suggested to high-risk no-show patients, taking into account the disease, location, and the patient's mobility.

In addition, **an overbooking strategy was established to reduce uncertainty on time slots that are more likely to be skipped.** As an example, the predictive analytics system highlighted Thursday and Friday mornings as time slots with the highest no-show rates. The scheduling system used precision overbookings to make sure staff's time was not being wasted.

The no-show rate is now down to 4%, resulting in an annual savings of $3 million.

CHAPTER 4

Creating a Proper Data Structure for a Complete Analytics Methodology

Data analytics in the healthcare industry has a bright future. This is due to a number of factors working together:

- *A huge amount of data (150+ exabytes as of 2012);*
- *A desperate need to understand raw data and assign meaning;*
- *The significant number of business-oriented applications to which data analysis can be applied in the healthcare sector.*

Right now we are only at the tip of the iceberg in terms of implementing data analysis in healthcare. Going forward, however, there are some cautionary steps that could easily trip-up any organization pursuing an analytics platform.

In order to successfully implement a predictive analytics solution in the healthcare industry, it is necessary to have a clear vision of outputs, implement IT systems that are interoperable, and have a commitment to knowledge sharing across the organization.

To Know Before you Go

Identifying Business Needs

The application of data analytics in the healthcare industry is designed to make a concrete impact on existing business processes and to improve efficiency. In most industries the data already exists and, in the healthcare sector, it is in abundance. The challenge has always been, "What do we do with all this data? How do we add meaning to it?" Data analytics provides pathways to answer these questions — your ultimate destination depends on your business needs.

> *Making all of this happen requires a comprehensive data analytics platform that is capable of not only handling, automating, and visualizing data, but can also be used as a collaborative tool for different user profiles (e.g., IT, business, marketing).*

Tackling the no-show issue is just one example of how data analytics can transform entire business segments; predictive analytic methodologies could be equally applied to physician profiling, precision medicine, disease management, and so on.

Making all of this happen requires a comprehensive data analytics platform that is capable of not only handling, automating, and visualizing data, but can also be used as a collaborative tool for different user profiles (e.g., IT, business, marketing).

Leveraging your Data

The "analytics" part of data analytics sometimes steals the show, but the hard work occurs during the early stages: collecting and cleansing the data. The first step on a data project is to define the inputs — where is the data coming from? After data sources are defined, we progress to data cleansing which accounts for more than 80% of a data scientist's work. **These tasks revolve around standardizing the dataset and dealing with issues such as missing data, redundant data, and unformatted data… all of which needs to be parsed and formatted.**

An advanced analytics platform should be able to automate all of the above tasks, effectively freeing up a significant portion of labor hours spent doing monotonous data cleansing work.

End-User Focus: Deploying to Existing Systems

All of the winning algorithms and awe-inspiring models in the world are useless if the end-results cannot be effectively deployed to the relevant business process or system. In the case of no-shows, it is critical that schedulers be able to easily access scoring data so that they can make meaningful time slot suggestions to high-risk no-show patients. **In a healthcare environment, analytics outputs have to be made available to those working in operations (e.g., nurses, aids, physicians, insurance analysts).**

It is therefore critical that an analytics platform should be highly collaborative, easy-to-use, and accessible. It should not be a tool for data scientist alone but, rather, an intuitive solution that can readily be used by those with both IT and non-IT backgrounds. Projects should be shareable between users and editable in a team-friendly interface.

Developing System Interoperability

As mentioned, **the healthcare industry is highly fragmented when it comes to data**. Every department seems to have its own standards and data entry tools. These standalone silos of knowledge pose a significant challenge to data analytics initiatives — data scientists face numerous obstacles in terms of securing access to multiple data sources and then standardizing that data after access is obtained.

System interoperability means that different datasets can be used, regardless of source. An advanced analytics platform should have the capability to connect to multiple dataset sources and effectively combine them with other internal, or external, datasets. In other words, there should be no obstacles in-place when it comes to accessing data.

Proprietary, or closed source, systems introduce a long host of limitations and restrictions, particularly when it comes to scalability and data connectivity. Never pursue a proprietary system and always ensure dataset interoperability.

Fostering the Distribution of Skills & Knowledge

All of those individual silos of knowledge represent more than actual datasets... they also represent micro-cultures within large organizations. The larger the organization is, the more likely it is that "This is how we do it!" attitudes are prevalent. Generally-speaking, **there is often a hesitation to adopt new approaches and solutions;** sometimes there is also a disinterest in sharing information across departmental lines. Even if there is an interest, it's more than likely that the current technology does not support data sharing.

Data from all sources should be made accessible to the analytics platform so that a single source of truth can be attained.

These limitations, whether human or IT-based, all represent barriers to knowledge sharing. A good analytics platform should act as a catalyst in terms of helping management to break down those barriers. Collaboration, content sharing, and dataset connectivity are all features that can help healthcare payers & providers implement data transparency across business segments. Data from all sources should be made accessible to the analytics platform so that a single source of truth can be attained. Ultimately, all parties benefit, particularly the patients.

A Shift from Retrospective to Prospective Analytics

Being data-driven is not an option, as nearly every healthcare setting has too much data to use effectively. The key is to transition from traditional retrospective analysis to the more forward-thinking prospective analysis.

The former tells you that there is an existing problem and delivers analytical content based on that problem — the latter predicts upcoming problems so that they can be anticipated and their effects mitigated.

Healthcare providers are reacting to no-shows instead of proactively addressing the reasons why they are occurring.

From a business standpoint, retrospective analysis provides high-level visual summaries while prospective analysis is highly focused on a single well-defined business problem.

Engaging with Patients: Now or Never

Data analysis has the capability to provide powerful insights into events that were, in a previous life, comprised of raw unformatted data. Sometimes connections can be made between datasets and data points that were unanticipated.

There is no reason why these insights should be limited to healthcare providers — **why not share relevant data insights with patients in order to engage and empower them?**

For example, patients with special mobility needs may enjoy seeing data visualizations that convey how transportation schedules are used to provide more relevant appointment schedules.

Conclusion: Curing the Healthcare Industry One Data Product at a Time

It's clear that when a patient does not appear for an appointment, both time and money are lost. **The issue has now reached a stage where the healthcare industry, as a whole, is losing billions of dollars each year.** Attempts to fix the problem are really stopgap measures designed to address the symptoms.

Data analytics enables organizations to stop the guesswork and understand exactly when specific patients are likely/unlikely to appear for any given time slot.

In fact, the core issue is being ignored completely: healthcare providers are reacting to no-shows instead of proactively addressing the reasons why they are occurring. This knee-jerk reactionary approach has resulted in policies that not only do not stop the financial loss, but cause needless patient discomfort, increased waiting times, and negative doctor-patient experiences. From charging fees for no-shows to cutting precious appointment times in half, these **misguided remedies have inadvertently created contentious doctor-patient relations instead of fostering amicable & friendly relationships.** Data analytics enables organizations to stop the guesswork and understand exactly when specific patients are likely/unlikely to appear for any given time slot.

At its core, predictive analytics cuts through, clarifies, and conveys highly relevant information based on a wide array of diverse data. Local data (e.g., patient information and historical results of appointments) is combined with global dimensional data (e.g., transportation costs, traffic routes, weather, geographical

distances, and patient diseases) to create a holistic view of the variables that are affecting no-show rates.
Models are created, tweaked, and refined until a clear picture emerges that explains why patients are not appearing and, more importantly, what your clinic or hospital can do to directly address the core problem.

The possibilities for predictive analytics are endless and are indicative of the world we live in: where vast quantities of raw data can be accessed, cleansed, collected, parsed, formatted, and elegantly visualized in a meaningful way.

Predictive analytics is rapidly changing the way business is done in the 21st century.
No-shows are a critical issue that have a negative impact industry-wide but, at the end of the day, it is a singular problem in a vast sea of possibilities. The reality is that the healthcare industry faces a plethora of challenges whose solutions revolve around vast amounts of untapped data.
What if patient satisfaction data could be correlated with healthcare fees? What if Internet of Things (IoT) sensor data in hospitals could be used to predict medical appliance needs? What if EHR and global data could be used to predict patient non-compliance with medications?

The possibilities for predictive analytics are endless and are indicative of the world we live in: where vast quantities of raw data can be accessed, cleansed, collected, parsed, formatted, and elegantly visualized in a meaningful way.

The future of predictive analytics in the healthcare industry is indeed bright and, whether the subject is no-show issues or a different challenge all together, we look forward to discussing the possibilities with you.

End Notes

1: Maalinii Vijayan, “No Shows: Effectiveness of Termination Policy and Review of Best Practices”, Wright State University, 06 November 2013, 5.

2: Weisner C, Mertens J, Tam T, Moore C., “Factors affecting the initiation of substance abuse treatment in managed care”, Addiction, 2001;96(5):705–716.

3: Mitchell AJ, Selmes T., “A comparative survey of missed initial and follow-up appointments to psychiatric specialties in the United Kingdom”, Psychiatric Services, 2007;58(6):868–871.

4: Gaus, Clif, “National ACO Survey”, National Association of ACOs, 21 January 2014, 3.

5: Mathematica Policy Research, “Health Information Technology in the United States, 2015: Transition to a Post-HITECH World”, Harvard School of Public Health, et al., 2015, 54.

6: Hoover, Waco, “Transforming Health Care Through Big Data”, Institute for Health Technology Transformation, 2013, 5.

7: Allen, Arthur, “Critics warn of $11 billion Pentagon health records fiasco”, Politico.com, 28 July 2015, retrieved 05 December 2015 from http://www.politico.com/story/2015/07/pentagon-electronic-health-record-critics-120730.

8: Manyika, James, et al, “Big data: The next frontier for innovation, competition, and productivity”, McKinsey Global Institute, May 2011, 11.

9: Manyika, James, et al, “Big data: The next frontier for innovation, competition, and productivity”, McKinsey Global Institute, May 2011, 10.

10: Toland, Bill, “No-shows cost health care system billions”, Pittsburgh Post-Gazette, 24 February 2013, retrieved 05 December 2015 from http://www.post-gazette.com/business/businessnews/2013/02/24/No-shows-cost-health-care-system-billions/stories/201302240381.

Acknowledgements

This ebook was made possible by the contribution of healthcare professionals, researchers, industry experts, and our own customers.

Dataiku would like to thank Dr Martin Pusic, Director of the Division of Learning Analytics for the Institute for Innovations in Medical Education at NYU School of Medicine. His views on data analytics tools were extremely educational and helped us better understand the stakes at hand. If you'd like to find out more, please read this interview, published on the 25th of November 2015.

We are also grateful to Dr William Tierney, President and CEO of the Regenstrief Institute. He guided us in grasping Electronic Health Records' limitations in the healthcare industry.

A special mention for Eric Kramer, our Healthcare Data Scientist and expert in the application of machine learning to discover genetic biomarkers. His precious insights and in-depth knowledge of the healthcare industry assisted us in positioning DSS as the perfect tool to help cure the healthcare system's data problem.

Finally, without Romain Doutriaux, Dataiku's Solution Marketing Manager, this ebook would simply not exist. Romain led all the research, interviews, and design initiatives involved in the project.

Last but not least, we would also like to thank our healthcare customers, both providers and payers, for their continuous support and the great work we managed to achieve together.

Healthcare is an information business where the difference between life and death is at stake. We firmly believe that predictive analytics have a key role to play in this regard and we hope you will join us soon in being a part of the solution.