

# When Healthcare Data Analysts Fulfill the Data Detective Role

By John Wadsworth

“A data detective uses data to blaze new trails to the Triple Aim. To understand how he does this, consider the engineering marvel that is a bridge. Its construction is preceded by much trailblazing and planning.”

The human mind can read words even when the letters are out of order. As long as the first and last letters are in the correct spot, the letters in between can be scrambled and the brain applies context to make sense out of the words. Artificial intelligence pioneer [Ray Kurzweil](#) believes that all learning results from massive hierarchical and recursive processes within the human brain. In the learning process of reading, we first recognize patterns of individual letters, then patterns of individual words, then groups of words together, then paragraphs, then chapters, then whole books.

Patterns in data, coupled with context, are important tools to the [improvement process](#). Discovering these patterns and putting them into context requires the special talent of a healthcare data analyst acting in the role of a data detective.

A data detective uses data to blaze new trails to the Triple Aim. To understand how he does this, consider the engineering marvel that is a bridge. Its construction is preceded by much trailblazing and planning. A bridge is an operational manifestation of somebody going before and proclaiming that a bridge is needed. Most bridges allow commuters to cross a path that was previously impassible. In a real way, a data detective allows the same thing. He helps identify the paths that are worth blazing and often supports the process of engineering solutions to consistently capture opportunities.

It pays to understand how a healthcare data analyst can take on the responsibilities of a data detective and how to grow this competency and support it within healthcare organizations.

## The Need for the Healthcare Data Analyst to Play Data Detective

Three cross-industry drivers have been impacting the need for data detectives.

**1.) There's been exponential growth in data capture.** The typical cellphone likely has less than eight gigabytes of

storage today, but in the 1980s, one GB was unthinkable. In 1981, Apple was selling [a gigabyte of storage for \\$700,000](#). Today, it's possible to purchase a terabyte of storage for \$10. This precipitous drop in the cost of storage, coupled with the exponential growth in processing power has made data capture very easy.

Healthcare has also experienced phenomenal growth in [data](#). Along with the clinical care provided, this industry now captures demographic data and what it costs in terms of dollars and skilled labor to deliver that care. It captures how patients feel about their care, and a host of ancillary data like income, comorbidities, family history, and genetics.

**2.) The second cross-industry driver is catastrophic events.** During a one-on-one meeting with Robert A. Smith, a 30-year veteran of the FBI, he spoke about how 9/11 dramatically influenced the role of analysts working at the Bureau. This change was precipitated by what happened after 9/11. A retrospective analysis of the data within the Bureau's systems revealed detailed information about a possible threat, but this threat wasn't detected because of two contributing factors: the FBI needed more analysts and they needed better analytic competencies. The problem wasn't with getting data, it was with interpreting the data and getting that interpretation to the right people. This resulted in a sweeping change in the Bureau and a tripling in size of the analytics team.

Unfortunately, healthcare also has catastrophic events, such as wrong-site surgeries, medication errors, and patient falls. Furthermore, the ever-increasing costs of care delivery, coupled with shrinking [reimbursements](#), force healthcare providers to do more with less.

**3.) Healthcare needs critical thinking around data.** The hype of [big data](#) and confusing messages around what technology can and cannot do is where critical thinkers thrive. Healthcare needs big data and platforms to provision big data and accelerate integration. But the problem with [healthcare analytics](#) today isn't that it needs more data, faster processing, or more sophisticated tools. Improvement in those areas will yield gains, but they will be marginal because they have outpaced the industry. The rate-limited factor today for healthcare analytics is the collective inability to think critically. Healthcare doesn't need big data as much as it needs critical thinking around data. Somewhere along

the journey, we started to believe an easy button existed somewhere, when it doesn't.

## What Is a Data Detective?

How does a healthcare data analyst perform the duties of a data detective and how does the role differ from related roles, like the department or business analyst? This is answered by looking at three levels of analytic complexity along with the technical skills and contextual understanding needed in each of these levels.

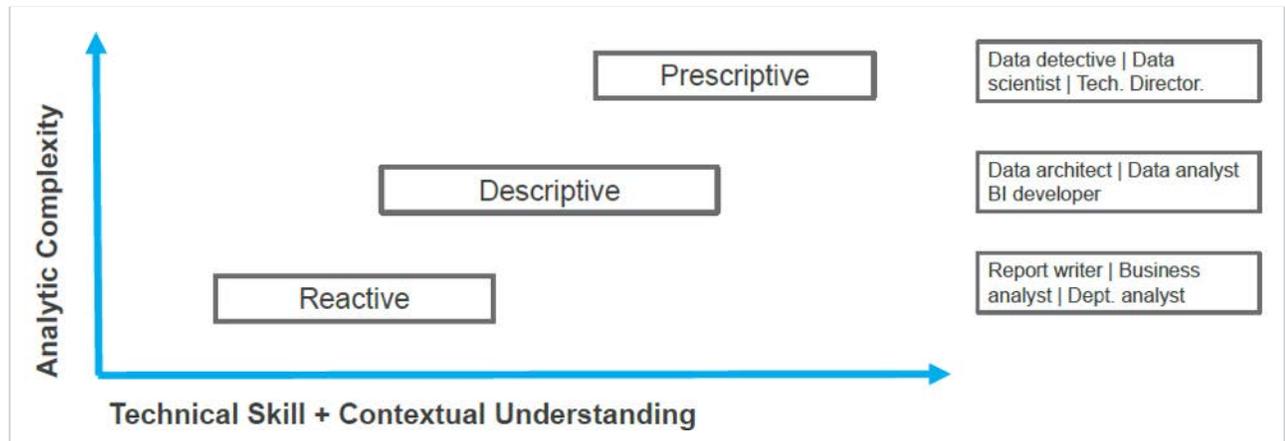


Figure 1: Appropriate roles for each level of analytics.

## Reactive Analytics

The first level is reactive analytics. These are managed by a report writer. Functionally, this type of analytics shows counts of activities or lists of patients and they include basic calculations on well-established, industry-accepted metrics such as diabetes admissions. Reactive analytics answer anticipated questions. A physician wants to know who has diabetes on her panel. A nurse supervisor wants to know how many hours of overtime ran through her department last month. Reactive analytics require minor tweaks to the look and feel of a report, but that is the only level of customization. Reactive analytics are generally confined to a single source of data and are often included as part of a vended system. Consequently, the report writer's contextual understanding of what is being measured and why it matters can be remarkably low. Reactive analytics explain some of *what* is happening, but they don't explain *why*. Report writers, business analysts, and department analysts are often tasked with this analytic function. It's a good entry point into healthcare analytics.

“Why are data modeling and provisioning so important? Descriptive analytics get much closer to addressing the why as well as the what, but they have a dependency on modeling real-world concepts.”

## Descriptive Analytics

The second level is [descriptive analytics](#), which are managed by the data architect. These are moderately complex and attempt to describe healthcare in the world around us. Analytic work of this nature is done outside of vended systems, usually in a dedicated system, like an [enterprise data warehouse](#) (EDW). These analytics leverage highly customizable data models that are populated with multiple sources of data from the EMR, hospital and professional billing, pharmacy, and labs. Data provisioning and integration efforts into the EDW afford a more comprehensive view of the activities within the health system as a whole. Work in this area is an ongoing and iterative process.

For example, a health system is motivated by penalties to reduce its heart failure [readmissions](#). With this focus, the cohort definition and readmission criteria are explicitly defined by CMS, leaving little room for interpretation. Descriptive analytics leverage that CMS construct to determine what gets loaded into the data model. The data architect works with nurses—subject matter experts—to scour approved data sources—diagnosis codes, admissions, discharge codes, CMS-defined patient types—to populate the model.

Why are data modeling and provisioning so important? Descriptive analytics get much closer to addressing the *why* as well as the *what*, but they have a dependency on modeling real-world concepts. Data models attempt to establish a comprehensive view of what is *supposed* to be happening (hence the data model), and what is *actually* happening. Descriptive analytics are often used to measure adherence to adoptive care pathways that have been vetted by clinical teams.

Data architects model what the care team is doing to the patient and how the patient fared as a result. Within this level of analytics, it's common practice to establish a [business intelligence](#) layer to consume portions of that data model with the objective of visually representing what is working along the care pathway and what warrants further attention. Descriptive analytics require significant interchange between technical experts and domain experts. As a result, these technical roles have a strong understanding of how analytics integrate into the workflow.

## Prescriptive Analytics

The third level is [prescriptive analytics](#), which are managed by the data detective. This type of analytics explains the *why*. Clinical improvement, waste reduction, and financial opportunities are first discovered using prescriptive analytics. Root cause analysis is a core function of this work, because once the processes that are contributing to waste are understood, it leads to the information needed to address meaningful change. These analytics are prescriptive because they clarify the needed action.

### A Tale of Two Domains

The value of prescriptive analytics is best illustrated with this data detective story. A pharmacist named Jack was working in a community hospital in the Midwest. The CMO, concerned with potential narcotics diversion, paid him a visit to ask how the hospital would fare in an audit. Controlled substance diversion was something Jack was passionate about, but he'd been unable to engage hospital leadership. That was until a recent news story about a competing hospital's narcotics diversion scandal had captured leadership's attention.

The CMO brought in Matt, a data detective, to work with Jack. The current process for detecting diversion was when a veteran pharmacist would be tipped off by observing erratic behavior in a staff member. The pharmacist would begin a lengthy review of nursing documentation on patient files within the EMR (a single data source). If anything seemed amiss, the pharmacist would sit with the staff member, along with nursing management and HR, to have a conversation. A single investigation consumed 45 hours of labor and spanned weeks.

Using this process, a case of diversion had never been detected within the hospital, even though benchmarking data showed that a hospital its size should have been turning up a case every 90 days. The hospital knew it had a problem; it simply lacked the ability to detect and quantify the magnitude.

Pairing Matt and Jack was the perfect combination of technical and domain expertise. Matt discovered data elements within multiple transaction systems, and each captured a different picture of the workflow involved in controlled substance administration. Matt pulled EMR data for every surgery and procedure that warranted a narcotic administration into the underlying data model he designed.

“The meaningful application of technical skills, coupled with the pharmacy domain expertise, allowed a manual investigation process to be automated. Automation allowed a daily review of every surgery, controlled substance administration, pain score, and nursing shift.”

“ A good data detective is self-aware enough to know when he doesn't know something. Technically, he is proficient at accessing, provisioning, data modeling, and analyzing data, but he often lacks meaningful context. ”

He pulled in patient-reported pain scores and coupled that with nursing shift data, pharmacy administration data, and data from the controlled substance dispensing cabinets on each of the floors.

Working closely with Jack, Matt was able to build an algorithm against the data model that highlighted statistical outliers in physician overrides by clinical support staff, and a frequency of spikes in patient-reported pain scores as entered by the nurses. Jack and the veteran pharmacist played key roles in validating the data and the algorithm.

The meaningful application of technical skills, coupled with the pharmacy domain expertise, allowed a manual investigation process to be automated. Automation allowed a daily review of every surgery, controlled substance administration, pain score, and nursing shift. The investigative work that previously took 45 hours over several weeks could now be done in five seconds. This freed up precious time from the highly skilled pharmacist and nurse managers. The hospital began to use the tool immediately. Outliers jumped off the page and in the first three weeks, three chronic diverters were identified.

### What Does a Data Detective Do?

Data detectives are story engineers. In a way, they give voice to the data. In the above story, Matt assembled narratives from data systems. He learned that orders for narcotics were placed within the EMR, but narcotics were released from locked cabinets that logged who actually accessed the cabinets and confirmed the physician order. Nursing shift data was kept in another isolated system that held time-tracking data.

A second function of the data detective is to work with other domain experts to bring together seemingly unrelated data. A good data detective is self-aware enough to know when he doesn't know something. Technically, he is proficient at accessing, provisioning, data modeling, and analyzing data, but he often lacks meaningful context. Armed with data profiling efforts, he partners with domain experts to collaboratively focus on the narrative and the white space around the narrative. Domain experts can quickly spot issues within captured data because they know what should and should not be in the data capture process. The combination of data narratives embedded in the workflow make things interesting. Matt tried to make chronological sense out of

surgeries, nursing shifts, pain scores, and narcotic order fills and overrides, but he couldn't do it without help. Bringing those data voices out of obscurity to the subject matter experts allowed them to connect the data meaningfully. From this example, finding spikes in pain scores was not unusual for some invasive procedures. Nurses captured this data as a routine part of patient recovery. However, finding correlations between pain spikes and nursing shifts told a different story than either narrative on its own. This correlation is an example of something found in the white space of the narratives.

A third thing that detectives do well is, using the same data, they make discoveries that others before them have missed. In the controlled substance diversion example, hundreds of people had access to the same system. They may not have been as technically savvy as Matt, but they had the same access.

This reinforces an earlier point: The problem with healthcare analytics isn't a lack of access to enough data. Healthcare systems don't have enough access to big thinkers of data.

### Three Ways that Data Detectives Go About Their Work

Data detectives go about their work by first asking a lot of questions, even to the point of annoyance. As noted earlier, they are some of the most self-aware people when it comes to understanding their own ignorance. Fundamentally, data detectives struggle with predetermined limits, especially those that are vendor imposed. Data detectives are frustrated when they see the inputs of a vended system. They will black box that data model, transform it, and then spit out reports. They want nothing more than to get into the black box and make sure it reflects workflow reality. If they can't, they will engineer their own construct.

Matt, the data detective, asked questions like, "Does every patient get a narcotic? If not, which patients do get them? If there is an order for a narcotic, is it always administered? If not, why not? Are all narcotics the same? Are some used for certain procedures, but not others?" The answers to these questions came from the domain expert, not the data person. Matt was trying to meaningfully represent in the data what was or wasn't happening in the workflow around narcotic administration. Questioning those involved became a two-way street between technical expertise and the workflow expertise. When those come together, they provide the right context.

Secondly, data detectives let the data inform. Effective detectives are good at keeping bias at bay. For them, this is a technical, logical, non-emotional exercise. As story engineers, they toss out multiple drafts of what the data could be telling them. They play with different storylines and look at different angles. Data detectives have the ability to *not* get married to an idea. As the novelist, [Stephen King](#), wrote in his book, [On Writing: A Memoir of the Craft](#), "... even though it breaks your egocentric, little scribbler's heart, kill your darlings."

Data detectives don't chase data in support of a foregone conclusion. Instead, they keep an open mind as to where the data, with all the combined narratives, can lead them. The data detective doesn't know the story that is going to be told at the outset. Matt's story began with a technical data profiling exercise. He paid attention to the completeness of the data, representative of that workflow from order to administration, and he paid attention to what was not in the data. For example, he was told that every cabinet was supposed to have required documentation. He found that many overrides didn't have this. He paid careful attention to the white space.

A third attribute of data detectives is that they know how to get to the heart of what matters. They understand the point to be made and how to get there quickly. Something is hidden in every set of data and data detectives go about exploring it to figure out the relationships, and then bring those to the forefront.

Making contextual connections can be taught, but some people have an innate ability to do this better than others. In the discovery process, there's an instantaneous shift in consciousness from confusion to understanding and data detectives do this very well. They find the relationships and, along with domain experts, figure out which ones are worth pursuing. When everyone else sees the relationship, they agree it's the path worth pursuing. Data detectives find the routes worth pursuing and determine the best place to build the bridge. Then data architects come in and operationalize those meaningful routes into bridges. After the bridge is built, report writers monitor the traffic that crosses it.

The technical work effects to support outcomes improvement can be distilled into three big work efforts:

A third attribute of data detectives is that they know how to get to the heart of what matters. They understand the point to be made and how to get there quickly. Something is hidden in every set of data and data detectives go about exploring it to figure out the relationships, and then bring those to the forefront.

**1.) Model real-world concepts in a database.** For example, when monitoring a diabetic population and representing that in a database, ICD codes, CPT codes, and hemoglobin A1c values can all be used to build a meaningful proxy representation of that construct in the data.

**2.) Analyze data to identify potential opportunities.** With the available sources within a system, give the most comprehensive view around that construct. Load it with EMR, claims, professional billing, hospital billing, and ancillary data.

**3.) Couple analysis with subject matter experts to find the needful action.** There's a strong correlation between sustained outcomes improvement and a permanent pairing of technical and domain experts. This combination enables organizations to identify and operationalize the paths worth pursuing.

### Supporting the Data Detective Career Track

The key skills and proficiencies to look for when recruiting data detectives involve SQL, ETL, data modeling, BI/ Visualization, and data analysis. These are required at varying levels within the three levels of analytics.



Figure 2: Skill levels of the five capabilities needed in the three analytics levels.

Seek out nerds with personalities. Health Catalyst finds plenty of candidates with the right technical skills, but they aren't the right fit. Only about three percent of candidates are hired. So what is it that makes technical staff so valuable to the organizations that they support? Health Catalyst conducted a cross-industry meta-analysis, looking up many technical roles, like data architect, data analyst, and data scientist. This went into a database to identify traits that are meaningful contributors to the value of each role. Then the traits were mapped into general categories and the highest-ranking category turned out to be “social,” not a description typically associated with technical people. But the traits that mapped into this category were team player, flexible, negotiator, facilitator, persuasive, and relationship oriented. Ranked second was the “smart” category, which captures the typical technical and IT skills. Ranked third was the “ever-learning” category, which includes traits like curious and inquisitive. Data detectives also care about the work they do and the impact their projects have on others. Technical skills are valued, and detectives can be super smart, do great analysis, and find awesome insights, but if nobody can work with them, then the impact of that analysis is limited.

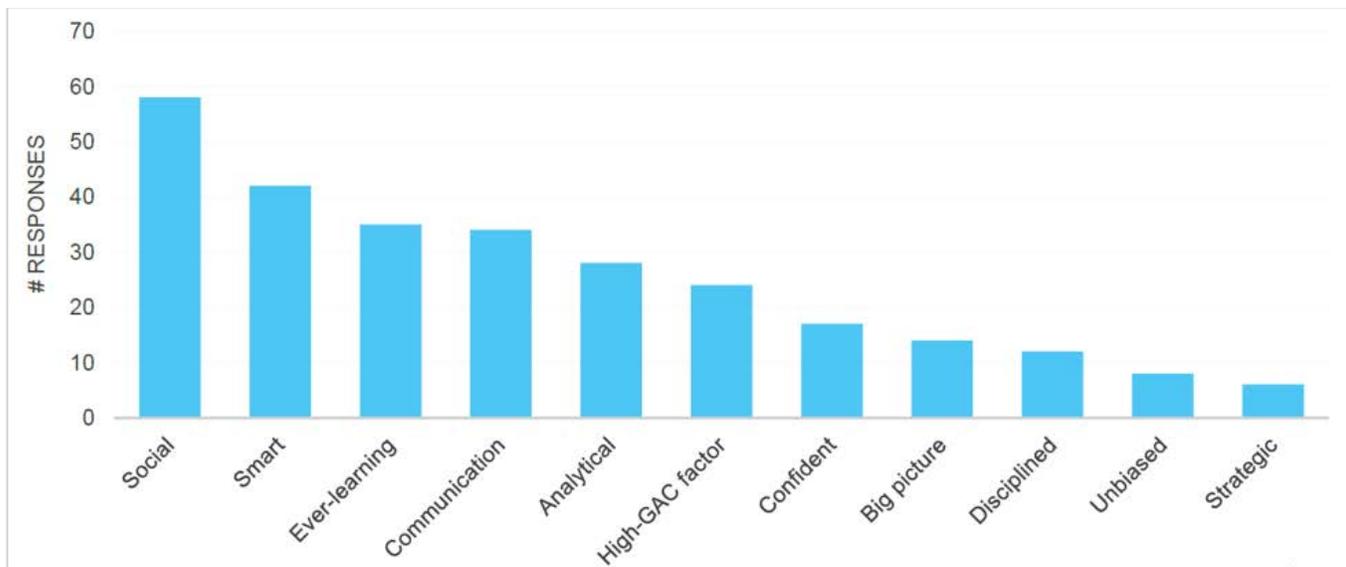


Figure 3: Attributes of a data detective.

## Develop the Data Detective Competency

Many organizations do not have a data detective role, but they diligently strive toward adding this valuable position. For the technical people pursuing this path, here are a few points of advice:

“The role of the data detective has never been more vital in healthcare than it is today. They are the big thinkers around data and a perfect fit in this big data environment. Prescriptive analytics, with their focus on reducing waste, optimizing financial opportunities, and improving clinical outcomes, require the inquisitive mind of a data detective.”

1. Improve your social and communication abilities with your consumer base. Pay attention to what is not being said. It will help you gain better appreciation for what happens to them. At the end of the day, if your report comes up short, find out why. Clinicians are the workflow experts who understand what's best for patients and your reports should reflect that.
2. Produce fewer, more meaningful reports that clarify needed action.
3. Analysis is king. Learn to think critically and give voice to your data.
4. The value of data detectives often manifests in “app-less” fashion. You don't need an app for everything.
5. Data detective skills can be learned on the job. The progression from report writer to data detective takes ~5+ years.

For those responsible for managing and growing data detectives:

1. Technology cannot stand by itself. It needs proficiency, which requires collaboration between IT and clinical, IT and operations, and IT and finance teams. Make this integration possible.
2. Request fewer, but more meaningful reports. Ask yourself why you need the data you're requesting.
3. Invest in your data detectives by prioritizing where the organization wants to focus. Establish a culture that elevates the analyst.
4. Recognize that growing data detectives takes time (~5+ years).

### Data Detectives: Valuable and Worth the Investment

The role of the data detective has never been more vital in healthcare than it is today. They are the big thinkers around data and a perfect fit in this big data environment. Prescriptive analytics, with their focus on reducing waste, optimizing financial opportunities, and improving clinical outcomes, require the inquisitive mind of a data detective. Data architects, BI developer, and report writers all uniquely contribute to analytic process, but healthcare data analysts, performing as data detectives, see all the connections, where others only see individual pieces to a puzzle.

Healthcare administrators should be consistently wondering what's missing from their outcomes improvement efforts. Specifically, analytics leaders should closely examine gaps in their strategies and capabilities and ask how they can fill them. The solution to these mysteries just might be found with a data detective.

### About the Author



John Wadsworth is the Technical Operations, VP at Health Catalyst. Prior to Health Catalyst, he worked for Intermountain Healthcare and for ARUP Laboratories as a data architect. John has a Master of Science degree in biomedical informatics from the University of Utah, School of Medicine.