A DAY IN THE LIFE OF A HUMAN DATA SCIENTIST:
Pursuing the Truth and Crafting Insights

Michael Kleinrock, September 2018
IQVIA Institute for Human Data Science

Making human data science a reality "takes a village." That village includes health science experts, IT infrastructure specialists, database administrators, computer programmers, coders, and statisticians. Their work comes to life when researchers and analysts translate de-identified "big data" into answers for some of the most important questions facing healthcare.

Michael Kleinrock is one such researcher. A nearly 20-year veteran of IQVIA, Michael serves as the Research Director for the IQVIA Institute for Human Data Science, bringing these transformative insights to healthcare stakeholders around the globe to drive deeper understanding and collaboration – all with the aim of improving human health.

So, what, exactly does being a human data scientist entail? We asked Michael for his views...

> *Human data scientists must think very hard about both the data they use and the implications of their findings on stakeholders such as patients and physicians.*
>
> *— Michael Kleinrock*
>   *IQVIA Institute for Human Data Science*

IT BEGINS WITH ASKING THE RIGHT QUESTIONS

I spend a majority of my time researching questions faced by my colleagues, clients or any one of the many different external stakeholders whom the IQVIA Institute supports as they prioritize how to use increasingly scarce healthcare resources. But I still have the latitude to step back and set the direction of my own research on broad, preliminary, even provocative questions that others haven't had time to think about:

- Are there gaps between the way the healthcare system is supposed to work and the way that it is actually working? Where are we falling short of the ideal?

- What are the current hot topics, and what seems to be missing from the debate?
- Where are the pressure points that are putting different stakeholders at odds with one another?
- Are we over relying on any research that is getting stale or too old given current conditions in the market?
- When I present a piece of research, what seems to leave the audience confused or eager for more detail?

In any given year, these questions lead to an exploration of multiple topics, which might range from the impact of drug pricing on consumption, demystifying net manufacturer revenues, explaining drivers of spending growth by brands or generics, exploring the emerging dynamics around biosimilars, to characterizing the clinical benefits of new drugs.

Despite the complexity inherent in these kinds of questions, a good data scientist, in my view, knows the value of simplicity in research. Although machine learning may be "sexy," it's not always necessary. Having the discipline to ask the right questions and determine the right methodology often means hearing hoofs and thinking horses, not zebras.

IT CAPITALIZES ON A VAST BODY OF COLLECTIVE KNOWLEDGE

I'm only able to tease out the meaning of our massive data sets because I stand on the shoulders of thousands of my predecessors and current IQVIA colleagues around the world. Together, we have collective knowledge drawn from decades of analyses. Plus, we all speak the same language, so we know how to interview one another about what we've learned in order to perpetuate our understanding. This might be as simple as knowing when there is a null versus a zero in a record, and whether that is interpreted

as a data gap or a true absence; a seemingly small, but critical distinction. And much of this is built into the IQVIA CORE, integrating domain expertise, data, technology and analytics that allows us to keep pushing our work forward.

Good data scientists are fully transparent about their work. They document what they did, why they did it, where the information came from, and what it revealed. But as with any documentation, it is only as useful as the shared technical language and experiences of those exchanging it. By developing practice areas and dedicated teams for the more common types of analytics, IQVIA has codified information and has created a supporting infrastructure that makes it possible for human data scientists to access and trust the data.

IT CONSIDERS THE HUMAN CONDITION

Data scientists – as opposed to human data scientists – search for answers without having to think past the finding. They can use brute-force programming to find correlations in a data set and be done with it.

Human data scientists, on the other hand, must think very hard about both the data they use and the implications of their findings on stakeholders such as patients and physicians. There are plenty of ways to use – and misuse – large healthcare data sets. If you don't know all of the nuances, you can get very tangled up in the details. At the same time, if you don't dig down a couple of layers to understand the root cause of a difference or change that's observable in the data, you can draw the wrong conclusion. Rarely do people dig deep enough into the evidence at hand. And all of this data must be handled to ensure privacy protection and correct governance. I work with non-identified data, but still adhere to all principles of privacy.

Brute-force programming and correlation alone will not be helpful or accurate in healthcare. Human data scientists expend a lot of effort trying to avoid mistakes and biases because we know the risks. And we know what we don't know. So, we tap clinical experts to help with market definitions, analytical experts to avoid false positives and negatives, and practice areas around each of the major analytical and business question areas.

Take, for example, a classic benchmark of market demand for prescription products: prescription volume. When this measure suddenly declined precipitously at the same time that the media were covering the high cost of prescription drugs, many people assumed that the decline equated to a drop in demand due to drug costs. In reality, the drop in prescription volume was due to pharmacy chains' new policy of dispensing 90-day supplies rather than 30-day supplies. That means that a three-month supply was counted as one prescription instead of three. So, although prescription volume dropped significantly in that channel, actual drug consumption had not changed.

IT DEMANDS FLEXIBILITY

Part of what I enjoy so much in my work comes in creating fresh approaches and in having several options for deriving an answer. Analyses that can only be done in one way – especially when the validity of the approach can only be known at the end – are to be avoided, if possible. Ideally, with each analysis, we build on the foundation of best practices. But then we improve upon them with each iteration. Our approaches must be both efficient and repeatable.

Even after having spent two decades at IQVIA, I'm still discovering new or unfamiliar data sets. The challenge in working with them is to know if you are able to extract the data you need on your own, or if you need to work with an expert who can extract data upon request. The more complex de-identified datasets are powerful tools but using them without any direction as a data mining exercise is unlikely to produce usable results. An optimal extract relies upon a market definition (which diagnoses or product codes are necessary), but it also relies upon some pre-definition of summary aggregation data attributes including disease groupings. Nobody wants to work with 17,000 ICD9 or ICD10 codes, but in order to be accurate and appropriate, we must define diabetes, or primary hypertension, or melanoma in a very specific set of codes. Even two decades into my career, I know I will always lack the expertise to

know the right codes for each market, and I can't be expert on seven database programming languages or platforms at the same time.

IT MEANS COMMITTING TO A JOURNEY

One challenge of this work is that no analysis is ever really done. This is a journey. Even when I've arrived at an answer, I have to consider it a working theory that has not yet been proven as "law" in scientific terms. I always have to be open to developing and sharing contrary evidence, if it exists.

IMS Health & Quintiles are now

IQVIA™