

Demystifying Data to Support Your Business Goals:

The Foundations of Healthcare Data

In today's business world, data—and the desire to capture insights from it—has become more mainstream. An Experian study found 83% of businesses rely on data as an “integral part of forming a business strategy,” while a recent IDC study predicts that by 2025 the amount of data produced will grow 10 times the amount in 2017. Companies now have the ability to capture data on just about everything. This increasing demand for more and more data has been supported by technological advancements, leading to an explosion in the breadth and depth of data. In many cases, this has led to collecting data first and figuring out how to use it later.

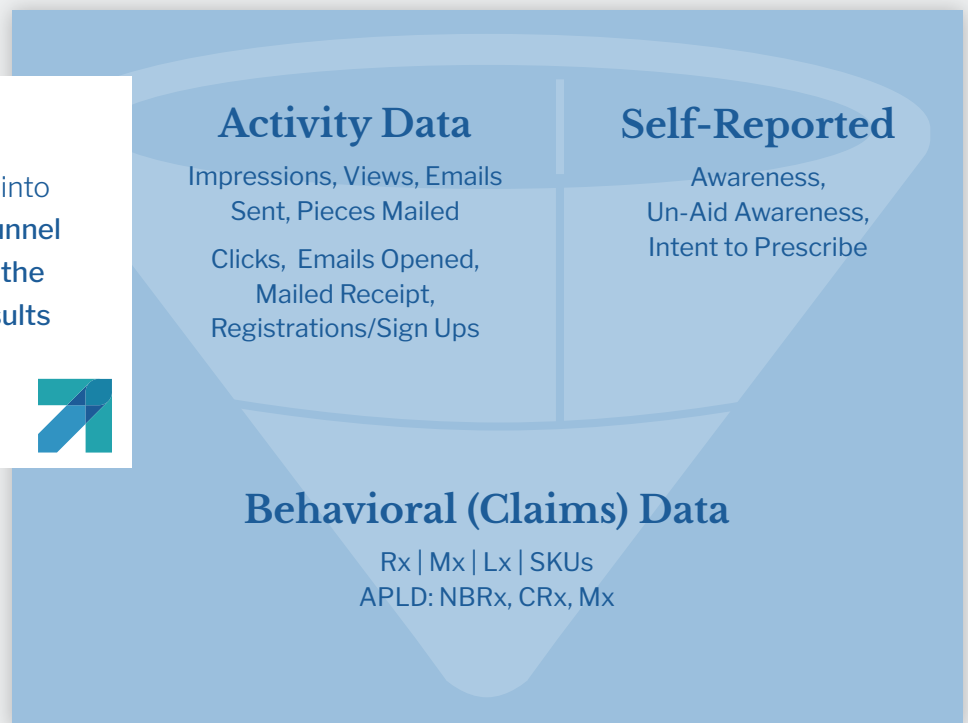
This blanket collecting of data combined with the desire for real-time information may not always produce positive results. Lost in the sheer quantity of data, we may become distracted from our end goals with metrics that are “red herrings.” Further, focusing on the wrong data can lead to misinformed decisions, which can negatively impact a business. The solution is to apply first principle reasoning, in which complicated problems are broken down into basic elements and then reassembled from the ground up. Using this reasoning, we are able to identify the metrics that align closest with our KPIs and focus on those.

This paper will explore the foundational sources and types of data, as well as the caveats that come with using each type.

Aligning Data to the Marketing Funnel

Applying a framework to data can help determine its appropriateness for business objectives. The first part of the framework is to overlay data onto the marketing funnel. Determining where in the funnel a metric or data source sits clarifies what insights can be supported. The key is to remember that data is a vehicle to get to an insight and data without insights has no value.

In the graphic, data is separated into two sections: **at the top of the funnel are engagement metrics and at the bottom of the funnel are the results of those engagements.**



While there can be plenty of discussion on the precise definition of top-of-funnel metrics, these are generally tracking the number of potential exposures or engagements, such as impressions and Gross Rating Point (GRP). With developments in how these metrics are captured, they have become less estimated and more precise; however, there are still potential exposures that we can confirm were served but viewership is implied. This is where the lower portion of the top of funnel comes into play. Clicks and opens are confirmation of engagement. Although there is still some wiggle room regarding human intent, these metrics take us a step closer to end-goal actions.

Types of Healthcare Data

Moving down the data funnel we shift from exposure and activity data to results data. Here, understanding the data types and, even more so, their appropriate application can get more complicated. At the highest level, there are three basic data types:

Direct Distribution Data (DDD), also referred to as “sell-in data.” This data tracks the movement of product(s) from the manufacturer or, more commonly, wholesaler to the pharmacy, hospital, physician office or system.

Claims Data or Anonymized Patient Level Data (APLD) is collected at the most granular level, stripping Personal Health Information (PHI) from pharmacy or medical claims through the adjudication process from middlemen in the payment chain called “switches” or from the “payer.” It is then reported at various levels from claims to physician and all the way to national.

Electronic Health Records (EHR), used interchangeably with Electronic Medical Records (EMR), is anonymized patient chart information. This rich data provides insights into things like disease stage or progression and recommended treatment(s) from diet and exercise to prescribing. The data is coming from physician notes via the EHR system and can be less structured than claims data. Although it provides insight into the physician’s recommendations, it does not always correlate to the actual behavior or treatment of the patient.

Of these data types, claims data is often the most common and appropriate data type for commercial analytics. There are two main categories from which this data is sourced:

Open-source data refers to data that is collected from a source below the payer. The benefit is that it includes claims from a range of payers and plans which contributes to its representativeness. The downside to open-source data is that patient capture can be fluid as they move in and out of the data set.

Closed-source data is collected from payers (i.e., insurer). It has the advantage of knowing when a patient enters and leaves the data set through member enrollment and cancellation. As a single data set, it also provides a complete picture of all adjudicated claims with the ability to precisely track a patient across hospital, laboratory, physician and pharmacy categories. However, this strength is also its weakness as closed-source data representation of the total market can be skewed by formulary and contracting. Thus, this data is more often used for health outcomes analysis rather than commercial analytics.

There are four main types of claims data: **Pharmacy, Medical, Hospital** and Laboratory.

We will focus on the most commonly used data sources: **pharmacy** and **medical** claims data.

Additional Data Descriptors

PHARMACY (Rx)

MEDICAL (Mx)

Data Source

HOSPITAL (Hx)

LABORATORY (Lx)

Pharmacy Data

Pharmacy data is commonly used in commercial analytics. This data is collected at various levels, from pharmacy to payer. With data now being collected at the patient level it provides two primary types of metrics: prescriptions and patient metrics. Prescription metrics are presented as NRxs (new prescriptions) and TRxs (total prescriptions), and patient metrics are often referred to as NBRxs (new patients) and CRxs (continuing patients). These notations refer to new to brand Rx and continuing Rx.

Understanding the difference between these metrics is critical as they provide different information. Starting with an NRx, this is a new prescription given to a patient (or sent to pharmacy) being treated for the first time or to a patient that has been on therapy already. In many cases, an NRx will have refills associated with it; therefore, the NRx and refills (RRx) make up a TRx (total prescriptions).

Shifting to patient metrics, this enables further segmentation of an Rx in order to determine the type of patient who is filling the prescriptions. There are three main patient cohorts:

- Naïve or New to Therapy, who are patients starting treatment for the first time*
- Switch/Adds, who are new to the brand of interest, having been or are continuing to be treated with another drug in the same market
- Continuing patients (CRx), who are patients that have already been treated with the drug of interest*

The first two of these cohorts, Naïve and Switch/Adds, are both considered New to Brand (NBRx).

The Anatomy of a Script

TOTAL RX

New Rx (NRx)			Refill Rx		
New to Brand Rx			Continuation on Brand Rx		
New Therapy Starts	Switch To	Add On	Continuation (NRx) Rx	Restart Same (NRx) Rx	Continuation (Refill) Rx

Now for the asterix.*



Patient level data is amazingly powerful, but it is very sensitive to business rules. With patient cohorts, lookback periods are selected to determine if a patient is an NBRx or CRx. The lookback period is a timeframe used to determine whether a patient has had a prescription for the drug of interest or one from the market basket of competitive products previously. Typically, lookback periods are set at six or 12 months depending on therapeutic class. Changing the lookback period and/or the market basket will change the NBRx and/or CRx value, meaning, unlike an NRx, an NBRx is not always an NBRx depending on the particular dataset pulled.

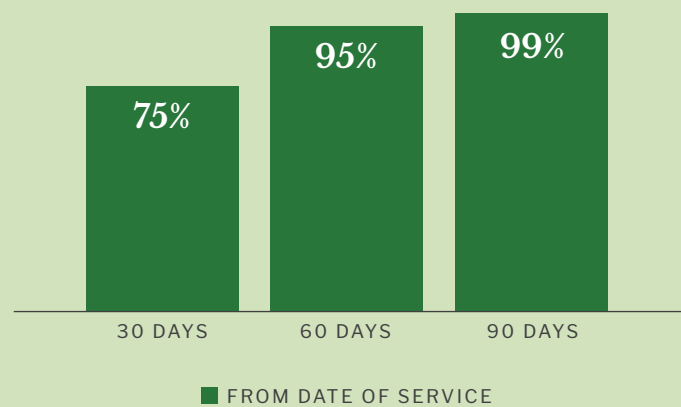
Medical Claims Data

The other commonly used data is medical claims. This includes office, long-term care and hospital inpatient and outpatient data. Though medical claims data include both inpatient and outpatient data, they are often treated as separate data sets because they come from different systems/forms (UB1500: Practitioner vs UB04: Hospital) and have different nuances. In most cases, medical claims data refers to outpatient office claims. Regardless, it is important to remember the source of this data is the billing process of a practice seeking reimbursement for services provided. It is also important to note that this data begins with a Medical Billing and Coding Specialist and then moves on to be reviewed by a claim's analyst on the payer end who may approve or reject the claim for payment. If a claim is rejected, the bill may be returned to the billing specialist for adjustment and re-submission. This process is one of the key issues with medical claims data: data lag.

Medical claims data lag means that approximately 75% of claims are available one month after the period end, and it may take up to 90 days from the end of the period to receive 95%+ of claims. **This means that at least a two-month lag is necessary for most analyses, as shown in the chart.**



% of Expected Data Available



There is a parallel that can be drawn between Rx and medical claims. A medical claim is a claim for reimbursement for a service provided. These services range from an exam, medical procedure or a procedure related to pharma, such as a vaccination or infusion. With the latter, a physician is billing for the administration of a vaccine or infuse product, which includes the cost of therapy. This means that the claim may or may not indicate the actual brand that was administered. In addition, a J-Code, which identifies an infuse product, may or may not be included in the claim(s) during a patient's therapy. This highlights the importance of working with partners who have a deep understanding of the data and how best to extract valuable and accurate insights.



Data without insights
has no value.



Conclusion

Data without insights has no value. Therefore, it is important we start with the insights we need and then identify the data that will help us uncover those insights. An informed understanding of the data and metrics enables marketers to select the most appropriate KPIs for their campaigns, which in turn influences measurement methodologies.

Look for future papers to come, in which we will dive deeper into aligning available data with the most appropriate campaign KPIs and methods to quantify performance.

¹ <https://blog.seagate.com/business/enormous-growth-in-data-is-coming-how-to-prepare-for-it-and-prosper-from-it/>

² <https://www.championbilling.com/breaking-down-cms-1500-vs-ub-04-formerly-known-as-cms-1450/> <https://images.app.goo.gl/CoGafXy35MYbiYU5A>
<https://www.verywellhealth.com/preparing-the-ub-04-form-2317061>

³ <https://www.archwayhealth.com/bundled-payments-blog/2015/09/bundled-payments-claims-lag>

⁴ <https://fs.blog/first-principles/#:~:text=Sometimes%20called%20E%80%9Creasoning%20from%20first,linear%20to%20non%2Dlinear%20results.>