



## MARKETER'S GUIDE SERIES

# NOT ALL DATA IS CREATED EQUAL

A Health Marketer's Guide to Selecting a Data Provider

### 3 PART SERIES:

6 Considerations for Selecting a Health Data Supplier

6 Considerations for Selecting an HCP Data Supplier

6 Considerations for Selecting a Consumer Data Supplier

# CONTENTS

---

## **Introduction** **3**

---

## **6 Considerations for Marketers When Selecting a Health Data Supplier** **4**

Type of Data: Deterministic or Probabilistic

Coverage: Portion of Universe Covered

Consent: Active and Direct

Aggregation: Linking Different Data at the Individual Level

Connected Data: Linking Cross-Individual Data

Time Lag: Immediacy of the Data

Summary

---

## **6 Considerations When Selecting an HCP Data Supplier** **9**

Type of NPI Data

Source of the HCP Data

NPI Validation and Verification

Informed Consent

Addressability

Accessibility

Summary

---

## **6 Considerations When Selecting a Consumer Data Supplier** **14**

Source of the Seed Data

Modeled Consumer Data Attributes

Targeted Consumer Audience

Consumer Data Seed Output Size

The Balance Between Scale and Accuracy

Consumer Data Model Maturity

Summary

# INTRODUCTION

Someone once said that data, not money, makes the world go round. Consumer data, now ubiquitous and sophisticated, has completely transformed how marketing and advertising get done. There isn't a single marketer today who isn't using data to drive media campaigns.

It's estimated that marketers in the US spend almost \$30 billion dollars on customer-related data. But marketers: Do you know what you're buying? Most buyers of data are focused on its usage, and rarely look into what's feeding the data they're using. And most data suppliers are selling extremely complicated, composite data products, so they're happy to keep it that way.

But this is unfortunate, because the fact is that data is not a commodity. Not all data is created equal. There are different sources and methodologies that go into building data products, and they make a difference in how you can use them and how well they perform. The more you know about what's feeding your marketing, the better you can manage your risk, control your output and results, and be smart with your spend.

There's another reason marketers need to become more knowledgeable about where the data they're using is coming from. While customers have thus far been willing to give away their personal information, there's now a trend among privacy advocates to take back control. And these advocates will start to hold marketers accountable for how they're using the data.

We invite you to walk through "Six Questions to Ask" when selecting a health data vendor, then drill into HCP data and DTC data considerations separately, depending on what you're looking for. Upon reviewing this content, you will have learned enough to make informed decisions about what data to buy, whom to get it from, and how to use it responsibly.



# 6 CONSIDERATIONS FOR MARKETERS WHEN SELECTING A HEALTH DATA SUPPLIER

The past decade has been great for health data. As a result of mandating EHRs via the HITECH Act in 2009, health marketers now have, at their disposal, an abundance of electronic (i.e., usable) health data, being provided by innumerable data suppliers. And being “data-driven” is now a non-negotiable marketing skill.

Yet, frustrations abound with regard to data-driven marketing. One of the drivers of these

frustrations lies upstream in the marketing process: the data itself. While health data sounds like something objective, a uniform widget that you should be able to buy, in reality *not all health data is created equal*.

If you're a marketer working with data to target media, personalize journeys or experiences, or measure campaigns, here's what you should consider when picking a data supplier.



## 1 TYPE OF DATA: DETERMINISTIC OR PROBABILISTIC

First, understand what type of data is being provided. Deterministic data, also referred to as first-party data, is actual information based on real users, and is known to be true. On the other hand, probabilistic data is derived data, and is based on relational patterns or statistically driven outcomes.

The validity of probabilistic data is predicated on two things: the size of the underlying database, and the accuracy of the models. In very large categories where we can identify big statistically significant samples, *and* where we know a lot about relational significance, probabilistic data can work. But modeled data sets don't work for smaller or newer categories, and even in more established conditions, you're going to see a fairly high margin of error.

In contrast, with deterministic data, you know you're targeting a real person. If the data provider has a solid verification process, you can be even more confident that your dollars are being used to reach or measure the person you're targeting.

## 2

### COVERAGE: PORTION OF UNIVERSE COVERED

Next, ask about coverage. The data supplier will be able to tell you what percent of your universe they can reach. Your goal should be to get as close to 100% coverage as possible, preferably with deterministic data. Higher coverage means you're more likely to reach all candidates in your target. Also, a wider data set of "real" records ensures that your measurement and feedback systems are more reliable.

Ask your supplier about their coverage. And find out about the balance of first-party vs. derived data they use to get to that coverage.

## 3

### CONSENT: ACTIVE AND DIRECT

Consent refers to the permission people give us to market to them. It originally came about to regulate email marketing (in fact email marketing used to be called permission-based marketing), but now, post GDPR and CCPA, applies to digital advertising as well.

There isn't only one way to get consent and not all consent is equal. The gold standard of consent is as follows:

- Consent should be *active*, where people actively *opt into* the program (vs. *passive* consent based on an *opt-out*).
- Consent should also be clear and comprehensive, meaning the person providing consent should understand what they are consenting to.
- Finally, consent should be specific to the organization that will be using the information, vs. applying broadly to "partners or affiliates."

When you're buying data, you're also buying the consent to use it. Find out what kind of consent your supplier has gathered.

# 4

## AGGREGATION: LINKING DIFFERENT DATA AT THE INDIVIDUAL LEVEL

A lot of health data is out there, but it's extremely fragmented. Prescription data is separate from lab and other reimbursement data. Socio-demo, retail, and "wearable" health data come from yet other sources. HCP data is not connected to the hospital the HCP works at.

That's why a question to ask your data provider is around aggregation of disparate data. How much information have they aggregated about each individual covered? And with what level of confidence, and confidentiality?

Confidentiality, or privacy, is such an important consideration with regard to health data that we're assuming it's implicit. But it gets complicated when aggregating data. If your supplier is aggregating data, they will need to demonstrate their ability to link records in a privacy-preserving de-identified way.

# 5

## CONNECTED DATA: LINKING CROSS-INDIVIDUAL DATA

If relevant, it's important to know if your data supplier has ways to connect non-parallel data sets, such as patient to physician or physician to hospital. This 360-degree view on the health decision process could enable you to design holistic push-pull brand experiences based on where the consumer or HCP is in their journey.

Not all data providers can do this. But it can be done.



## 6

# TIME LAG: IMMEDIACY OF THE DATA

Any data is available if you wait long enough for it to be scrubbed, cleaned, and combined into the right usable formats. But the usefulness of the data diminishes logarithmically with time, while the margin of error goes up.

The sooner you receive data, like when a person read specific content, saw your ad or downloaded your coupon, the sooner you can respond with personalized messaging. Similarly, with results, the quicker you know which combination of messaging drives real-world impact, the more responsible you can be with your spend.

Find out what your data supplier's time lag is for their data. Three months is way too long. So is one month. With automation progressing at such a rapid rate, you should be able to get close to real-time data and data updates, in an actionable format.



# SUMMARY

## WHEN MAKING A DETERMINATION ABOUT USING A HEALTH DATA PROVIDER, ASK THEM THE FOLLOWING QUESTIONS:

- 1 How much of your data is deterministic?
- 2 What's your coverage of my universe?
- 3 How active and explicit is your consent?
- 4 How much do you know about each person?  
Is it privacy-preserving?
- 5 Can you connect person-level data across groups?
- 6 How quickly can you turn around data and updates?

---

PulsePoint is built upon a foundation of first-party data, giving us the most accurate and robust point of view on health audiences. We're the largest fully-verified health audience solution, built using real-time behavioral signals, at scale. Health marketers gain intelligent insights on over 300MM health consumers and over 3.2MM HCPs to automate actions across devices for HCP and patient journey orchestration.



# 6 CONSIDERATIONS WHEN SELECTING AN HCP DATA SUPPLIER

Thanks to unique NPI numbers, every healthcare professional (HCP) in the U.S. is individually identifiable. As a result, NPI-level HCP data is abundantly available to health marketers, with suppliers competing mostly on comprehensiveness of coverage (i.e. percent of universe covered) and robustness of online and offline data variables associated with the NPI (i.e. how much do I know about the professional).

But differences in HCP data go beyond scale and scope. Data suppliers vary widely on type, quality, and usability of their data. To understand where your data supplier stands, here are some areas to dig deeper, and questions to ask:

## 1

### TYPE OF NPI DATA

The NPI, or National Provider Identifier, number is a 10-digit numerical identifier used to identify an individual provider or a health care entity. It is managed by the CMS, and is used for Medicare / Medicaid claims processing. It doesn't change with time, so it's a fairly failsafe way to identify each individual healthcare professional.

But what most people don't know is that there are two types of healthcare providers in terms of NPIs, Type 1 and Type 2. Individual providers get a Type 1 NPI, while practices or organizations are given a Type 2 NPI. An organization may have subparts that each have their own NPI.

Which NPI you should use to identify and reach your physicians depends on whom you represent and/or what you're trying to communicate. Type 2 NPIs are more relevant for payor-related products or services because, being at the organizational- or payee-level, they give you an accurate representation of practice level claims and practice patterns. If you're a product manufacturer and need to understand individual-level physician preferences and behaviors, Type 1 NPIs will be more useful.



## 2

## SOURCE OF THE HCP DATA

Data suppliers build their HCP databases and associated opt-ins by pulling from several different sources. They may use physician surveys, email marketing, publisher data, and/or may even scrape social media or public databases to append to the data they get from the NPI registry or the AMA.

While an amalgamation of data sources is expected, your rule of thumb should be that the closer the sources are to the HCPs themselves, the more reliable the data.

Dig into the data sources to understand where your data supplier gets their HCP data. Your goal should be to get person-level opt-in data from as direct and as close to the physician as possible.

## 3

## NPI VALIDATION AND VERIFICATION

While NPI numbers don't change, physicians move. They change practices. They add skills and certifications. Some even change specialties. Or they may decide to opt out on any given day. These changes are not automatically reflected in the NPI database. So how does your data supplier keep up with these changes?

Data suppliers have multiple verification approaches available to them, such as proactive outreach to the physician or practice to make sure the information is accurate, active usage of the data to ensure validity, or triangulation with other data sources.

You need to be confident that the data you're buying is accurate, up to date and has the appropriate opt-in for your intended use. Ask the data supplier if their data is verified, how, and how often. And beware of just one verification step: It's rarely enough to catch all updates across all covered HCPs. You should be looking for multiple sources of verification, ideally updated every 24 hours.

## 4

# INFORMED CONSENT

Examine the opt-in consent attached to the HCP database to find out how “informed” the consent is. What’s the level of knowledge of the provider:

- Does the provider know that they are giving consent?
- Do they know what they are opting into, i.e. what their data is going to be used for?
- Do they know who is going to be using their data?

While not legally required in many states, a best practice for consent is that the opt-in process explicitly names every organization that will be using the data, what data is being collected, and in clear, simple terms, every purpose they will be using the data for.

## 5

# ADDRESSABILITY

Having accurate data isn’t enough. It needs to be usable, and yield results.

Let’s say a physician is associated with the email address info@practice.com. You know who the physician is, you know their practice, and you have accurately identified an email address that’s actually associated with the practice. But in this example, it’s a generic email which the physician rarely, if ever, checks.

Or let’s say you reach the physician but he / she does not, and has never, responded to your messages. The result: higher campaign waste, lower performance results and a bad customer experience.

Make sure you’re buying addressable data: that you’re reaching real, known and active HCPs, and that you know enough about them to craft relevant experiences that are mutually beneficial.

# 6

## ACCESSIBILITY

Make sure the HCP data is available on the platforms you're looking to activate the data from, and in a format accessible for your marketing goal.

For example, if you are looking to simply activate a HCP segment in a DSP without the NPI-level attribution, an aggregated segment of cookies and device IDs would suffice. But if you are planning on running 1:1 NPI level reporting, make sure your data supplier can provide the appropriate opt-in, scale, onboarding, matching and reporting to support your attribution needs.



# SUMMARY

## HERE ARE QUESTIONS YOU SHOULD BE ASKING YOUR HCP DATA PROVIDER:

- 1 Where does your data come from?
- 2 How well are you matched to Type (1 or 2) NPIs to reach my universe?
- 3 What's your verification process and schedule?
- 4 How informed are the HCPs about the opt-in they're providing?
- 5 How addressable is your data?
- 6 How accessible is your data?

---

PulsePoint is the only activation and insights platform with its own authenticated, first-party dataset. Our solutions are powered by a database of over 3.2M verified HCPs with active and named opt-in consent, resulting in higher quality, more accurate targeting, and enabling more immediate insights than other platforms. Data sets are refreshed every 24 hours to account for any physicians opting out.

# 6 CONSIDERATIONS WHEN SELECTING A CONSUMER DATA SUPPLIER

Consumer data is big business. If you're a health marketer, you're probably buying or working with some kind of consumer health data. But NAI (Network Advertising Initiative) and DAA (Digital Advertising Alliance) restrict the use of personally identifiable health information (PII), except in cases where you have consumer consent. Given how difficult it is to get informed consent from large volumes of people, most consumer data that we buy is modeled data.

Because of how actively we've been using modeled datasets, they've gotten pretty sophisticated. But caution: not all audience models are the same. The next time you're looking to buy condition audiences or other health segments, here are a few things you should consider.

## 1

### SOURCE OF THE SEED DATA

Models are developed based on seed data. Data providers study a group of 'seed' people, then project out their observed attributes and behaviors to identify similar people in the larger population. It follows that the bigger and more robust the seed data, the more reliable the model. So make sure your seed data source is coming from national organizations or initiatives.

Data suppliers use multiple seed sources to get to opt-in health data, including surveys, coupons, offers, etc. All sources have their positives and negatives. Take surveys, for example. With surveys, the supplier gets opt-in consent, as well as behavioral and attitudinal insights into the survey respondent. On the flip side, these are self-reported so probably not completely reliable.

When purchasing data, make sure you understand how robust the seed data is and where it's coming from. Think about how accurate the seed data is, and how much it reveals about your audience.



## 2

## MODELED CONSUMER DATA ATTRIBUTES

Some data suppliers use medical claims data to seed their databases. These seed data are extremely robust and include real historical diagnosis, treatment and comorbidity information.

But since we cannot target using healthcare data, they model out this data to find other proxy targeting attributes like gender, age and zip code. So, if, for example, the model finds that high income women have a higher propensity to be diagnosed with Alzheimer's Disease (AD), then an AD drug using this model would end up targeting high income women with their ads. Clearly this approach yields innumerable false positives and false negatives.

It's important to understand how the model you're buying works. Ask the data provider what their data is modeled against, and what it's modeled towards. Get a sense of false positives and negatives, and model efficiency. And find out if and how they validate model output.

## 3

## TARGETED CONSUMER AUDIENCE

Does the data you're buying match the audience you want to reach?

Take contextual data for example. Let's say the segment you're interested in includes people who've consumed diabetes content. They are most likely to be motivated content seekers: people who are symptomatic but not yet (or newly) diagnosed, people who're having an issue or complication, or caregivers. Conversely, people who are living with diabetes long-term are probably not actively consuming diabetes content.

If you're looking for the total diabetes population, contextual data alone is unlikely to get you there. But if you're a drug or a device looking for the newly diagnosed, you can be confident that contextual data will get you to the right people. When evaluating an audience segment, try to get to the bottom of who is in the segment so you can evaluate if they're your audience.



## 4

### CONSUMER DATA SEED OUTPUT SIZE

The more you model out the seed data, the less accurate the output is going to be.

Here's how models work. Let's say you start with migraine sufferers in your seed data. You then use machine learning to understand the various attributes shared by these people. You create the model based on these attributes, then use the model to go find people who look like your seed people.

The first group of people identified are going to have extreme fidelity to the model. But this decreases as you bring more and even more people into the segment. At really high numbers, the model can get so diluted that it's worthless.

When working with data vendors, find out their seed to output ratio. If they're using, say, a seed group of 1000 people to create a segment of 10 million people, you know you have a problem.

## 5

### THE BALANCE BETWEEN SCALE AND ACCURACY

There's a constant tradeoff in the world of data between scale and accuracy. As discussed in the seed output size consideration, the bigger the scale, the lower the accuracy. When buying data, you need to decide where you want to land on that spectrum, based on your objective.

If you're launching, say, a disease awareness campaign, you want your message broadcast far and wide. In that case, opt for scale. However, if you're implementing a bottom-of-the-funnel conversion campaign, you want accuracy to ensure an efficient time-to- and cost-per- conversion.

## 6

### CONSUMER DATA MODEL MATURITY

Maturity is either a good or bad thing, depending on how the model works. Some models are static and get stale. Others, that rely on machine learning, get smarter with time. When making your decision about an audience segment, find out how old the model it's using is. Then find out if the model is blooming or atrophying with age.

In addition to these considerations, make sure you're familiar with the NAI guidance, so that you're keeping your organization in compliance.



# SUMMARY

Before you sign the contract with your consumer data supplier, you should ask the right questions about how the supplier is procuring their data. Not only will this help ensure you remain in compliance (by using data that was collected according to NAI guidance), but that your marketing efforts will have a higher chance of success.

**Consider asking the following:**

- 1 What's the source of your seed data?
- 2 What targeting attributes does your model use and how do you validate model output?
- 3 What target does the data cover?
- 4 What's the seed to output size ratio?
- 5 Where do you fall on the scale vs. accuracy spectrum?
- 6 How old is your model and is it getting smarter with time?

---

PulsePoint, the largest programmatic health consumer audience solution in the world, has the most robust first-party data on 300MM health consumers, giving us the ability to identify and reach patients at different stages of the health journey – in real time and at scale. Our data is addressable and actionable on day 1.

## ABOUT PULSEPOINT

PulsePoint is a leading technology company that uses real-world data in real-time to optimize campaign performance and revolutionize health decision-making. Leveraging proprietary datasets and methodology, PulsePoint targets healthcare professionals and patients with an unprecedented level of accuracy—delivering unparalleled results to the clients we serve. The company is now a part of Internet Brands, a KKR portfolio company and owner of WebMD Health Corp. For more information, visit [pulsepoint.com](https://pulsepoint.com).

To learn more about our data and technology solutions, and how they can support your business and its goals, contact us at [sales@pulsepoint.com](mailto:sales@pulsepoint.com).

**PULSEPOINT**<sup>®</sup>  
— LEADING HEALTH FORWARD