CentraForce Health Data FAQ

What makes the CentraForce Health dataset unique from other datasets that we can acquire?

There are a few things in combination:

First, our dataset is comprised of multiple, credible third-party consumer survey sources. These sources include locally sampled population surveys across healthcare and consumer measures that have been combined into a single questionnaire. Our licensor business partners include established research firms like Nielsen, Kantar, Symphony and many others. The healthcare/consumer surveys are updated two times a year and are locally sampled in all U.S. markets—210 in total. In the aggregate, our license sources include 150,000 individual measurements. Of these, we have curated more than 9,000 measurements for healthcare sector purposes and developed hundreds of composite measurements that will support many kinds of preventative, healthcare management and revenue optimization use cases.

Second, we purchase licenses from all sources at the de-identified respondent level which enables us to generate profile-able samples of persons who have self-reported their conditions, health behaviors, utilization patterns, health attitudes, marketing behaviors and much more. Our dataset provides an actionable view into virtually all aspects of the consumer's existence.

Third, we have developed the intellectual property that enables us to reliably project all of our consumer health intelligence onto a database of all persons and households in the U.S., including matching to de-identified patient records through a HIPAA-compliant process. What's particularly unique about our IP is that our projections are based on actual persons from a given market who self-reported to be in the defined population or cohort. There are firms who combine person-matched demographics with geographic datasets. The person-level demographic data has some value, but the behavioral and health insights are projected onto patient records from broader geographies and *are not sourced from persons* who are representative of the patient file or cohort population.

Fourth, our PopulationCentric Intelligence Platform includes a number of publicly available and privately licensed spatial datasets.

- Locations of healthcare professionals and locations of every type of healthcare facility.
- Database of all persons and households in the U.S. which includes detailed demographics and house level (privacy compliant) micro-analytics.
- Location and name of mapped destinations in the U.S., (e.g., all grocers, restaurants, bars, parks, fitness centers, points of transportation) categorized by type of locale.
- 40,000+ Census measures for every block group in the U.S.
- Relevant public health spatial data, such as crime, walkability and environmental risk.

What is a licensed data product?

A data product is a set of population measurements and data points that can be integrated into the client's data infrastructure and other workflows.

What purposes do your data products serve?

Our products are specifically designed to provide insights on socio-behavioral and engagement aspects of a population, cohort or de-identified patient record.

With these insights, our clients know:

- Whom to target
- What to do about it
- How to reach and engage them
- How to change them

The insights are applicable across use cases related to:

- Market analytics
- Medicare Advantage and other plan enrollment growth
- Socio-behavioral-driven utilizations
- Out-of-network cost
- Population health (patient- and cohort-specific preventions)
- Care coordinator outcomes
- Service line growth
- Social diagnostic tools that can be integrated to EMR

What kind of data are added to database records?

CentraForce Health data products include proprietary derivative measurements, composite measurements, modeled-weighted scores and location-based intelligence. These may be insights provided at geographic levels, such as county, zip code, block group or appended to de-identified patient records.

Are you brokering or aggregating data from other sources?

No. We license multiple, credible third-party data sources for the purposes of integrating and synthesizing them into equalized population-specific data points. Third-party data sources are not designed to work in harmony with each other or other data sources. CentraForce Health IP uniquely brings these different sources together in a way that is efficacious and representative of any given defined population.

What are the data fields contained in a data product?

At the individual or geographic record level, our data products can contain the following fields of information for both the population and its comparison populations:

- Measure Question
- Measure Answer
- Measure Category
- Measure Sub-Category
- Measure Topic
- Measure Sub-Topic
- Percentage of Population
- Population Number
- Likelihood Index (propensity)
- Weighted or Modeled Score
- Applicable Location-Based Datasets

How can I be confident that the data are accurate?

CentraForce Health has innovated micro-analytic sampling capabilities, data efficacy and testing processes that ensure data results are representative of any defined population. We have millions of de-identified respondents within our database who have responded to tens of thousands of survey questions. Our process allows us to isolate and generate profile measures for any defined “look-a-likes” within our respondent-level database.

For example, if the market is Seattle and we are profiling persons with Heart Disease, then we are sampling respondents within Seattle who have stated they have been professionally diagnosed with this condition.

How has this capability been validated?

CentraForce Health has its roots in over two decades and several hundred use cases within the healthcare and non-healthcare space. Over the last three years, CentraForce Health has validated use cases within the provider, payer, technology and life science sectors.

Is the self-reported data tested?

The original source data are from different licensors who carefully balance their samples to be reflective of surveyed population. We access reports that document and disclose how surveys were conducted and the results. While self-reported data reflect the self-perceptions of the respondent, there are a number of inherent controls, post survey, that help to ensure the respondents are honest in their answers.

Are smaller, or more remote, geographies able to utilize your process?

We start by making sure that our projections are based on actual people (respondents) who self-report to have the characteristics we want them to have versus making assumptions that everyone in the neighborhood looks alike. So, instead of basing our data on what is true of a neighborhood (top-down approach), we start with what is true of the specific kinds of people in any given market and then roll that up to the geography. Through this process, we are able to profile smaller and more remote geographies.

How old is your data?

Our data for any given deliverable is usually six months old.

Why does CFH use the term socio-behavioral risks of health?

Research has shown that social determinants are not enough and that health outcomes are also driven by behaviors. The socio-behavioral approach identifies the “manageable” behaviors, “shapeable” attitudes, and even “preventable” life circumstances that contribute to risk in addition to social determinants.

How are your different kinds of data products attributed to the individual de-identified patient record?

We use our same bottom-up methodology when attributing data product intelligence to the de-identified patient record. Our micro-analytics process cannot identify any individual person, but the deep intelligence that is known of actual persons can be leveraged. Because our respondent-level data have the same micro-data points as appended to de-identified patient records, we can create equalization between these two data sources.

What are your quality controls to ensure that the patient-attributed results are accurate?

In addition to the household matching micro-analytics, we are supplied with a HIPAA-compliant report of the patient record database that shows breakdowns across certain demographics, payer mix and health status variables. We carefully select from our pool of survey respondents so that they are consistent with the “known” attributes of the overall patient file.

What is the clinical basis for your socio-behavioral risk scoring system?

Our socio-behavioral risk scoring system was developed by population health clinicians.

How is socio-behavioral risk expressed…what do we learn from the data?

Each risk category is given a score based on weighted calculations around the degree of each variable's urgency, degree of practicality and degree of cost. There are two other key metrics included. The first is a frequency metric that tells you how many people (out of 10) are projected to exhibit the measured risk. For example, 7 out of 10 of the Asthma Population (or the database records) are likely to be at Risk for Depression. The second metric relates to likelihood, providing an index score based on how the population compares with another population. For example, if the Asthma Population (or the database records) have a likelihood index of 300 for Risk for Depression, then the Asthma population is 200% more likely than the comparison population to have this risk.

What is your engagement scoring system based upon?

Our respondent records report the degree to which they are engaged with any given channel, such as taking advice from their doctors, relying on collateral that they receive from a doctor or receiving a personal text message. They self-report how much time they spend with each channel, whether they value that channel as a source of healthcare information and whether they are likely to act after exposure through the channel. In addition to these three quantitative/qualitative dimensions, we provide channel usage insights that will inform outreach planning.

How would you compare your engagement data to those data sources that we can purchase from marketing data suppliers?

Our engagement data include measurements specific to the healthcare channel and healthcare content. Because we license data from multiple sources, we can provide a more comprehensive menu of engagement channels and channel usage detail. Another key distinction is that our engagement data is applicable to 100% of the population, whereas purchased marketing data applicability will range from 1% to 60%. Randomly matched marketing data are not collected in a way that is balanced or projectable onto larger (unmatched) populations.

How do you deliver your products?

- Data feeds
- Data interpretive reports or visualized insights, including use case-specific reports, infographics, dashboards