Veeva Crossix Data Platform and Methodology White Paper

May 2024

Introduction

The Veeva Crossix Measurement Suite enables marketers to maximize the impact of their marketing investments by connecting marketing to business outcomes such as new patient starts and incremental prescriptions (Rx). The platform uses real health data in a privacy safe way, fully compliant with Health Insurance Portability and Accountability Act (HIPAA) standards. All Crossix solutions share a common data model, which allows for combined analytics across media channels measured through Crossix as well as ingestion of data from Veeva CRM. Transparency is built into the platform, covering every stage from health and media data coverage, feasibility, mapping data for analysis, and sample sizes and waterfalls.

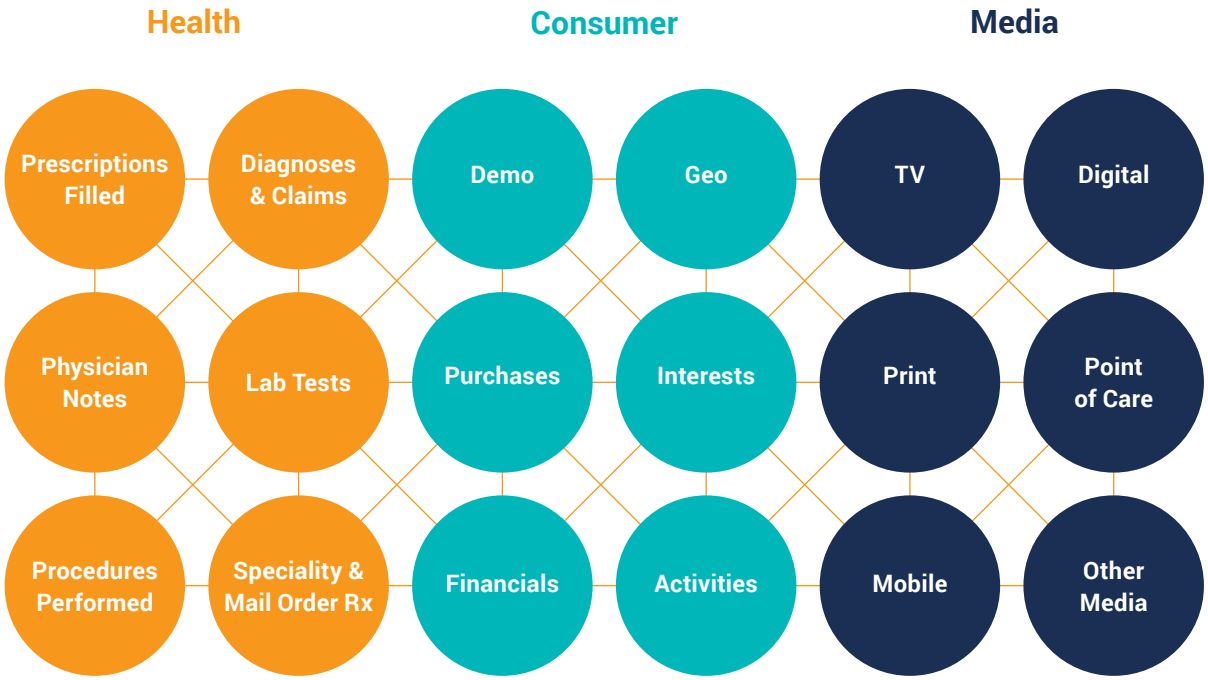
> *OBJECTIVE: This white paper provides an in-depth view of the Crossix Data Platform and Crossix Analytics Methodologies that power the Crossix Measurement Suite. This includes details on coverage of US health data and media exposure data, as well as the methodologies used to generate actionable insights for marketing optimization.*

Data Platform

The Crossix Data Platform

The Crossix Data Platform is the core of Crossix analytics and the only platform that can make health, consumer and media connections across 300MM+ patients. The Crossix Data Platform is built leveraging relationships directly with health data suppliers where we install patented SafeMine technology directly within the covered entity.

Our modern approach to analytics uses technology to combine large-scale data sets in a privacy-safe way. The Crossix Data Platform includes Rx, OTC, clinical, claims, consumer, hospital, media data and more.



Health data on the Crossix Data Platform consists of a unique set of prescription claims, medical claims, and EHR data sources that overlap to deliver more complete data – ideal for modern, complex therapies. It covers all distribution channels and therapeutic areas. This includes a combination of open and closed claim sources that captures a more complete patient journey. The unique set of data available in the Crossix Data Platform provides visibility into blocked products and Health Care Organizations (HCOs) that are not available from other sources.

HEALTH DATA SOURCES	
Rx data • Retail pharmacies (chains and independent) • Specialty pharmacies • Plan/PBMs • Switch companies / clearinghouses • Software providers	**Medical claims data** • Plan/PBMs • Clearinghouses • Software providers **Clinical data from EMR platforms**

Health data is updated daily. There is no lag for Veeva Crossix to access the underlying health data of the partners in the platform. As soon as Rx or Medical Claims are adjudicated, they are available to Veeva Crossix the following day.

The Crossix Data Platform is powered by SafeMine, a patented and proprietary technology for patient data linkage. SafeMine takes a unique approach to patient identity matching by linking records first and then de-identifying. SafeMine compliantly links patient records by using protected health information (PHI) and the full patient record over time, unlike the one record at a time used by conventional approaches. SafeMine linking securely occurs behind the data supplier's firewall.

SafeMine's probabilistic match algorithm is the only market solution that is accurate and robust enough to handle the rapidly growing volume of patient-level data in the U.S. and minimize both false positive and false negative matches at scale.

SafeMine delivers a superior approach to privacy-preserving patient record linkage. Patient data linked using SafeMine will deliver more accurate and complete insights and help optimize the business decisions that are powered by patient data. For more details on SafeMine, please see the Veeva Technical White Paper: Patient Data Linkage: Maximizing Privacy, Quality, and Accuracy with Veeva SafeMine.

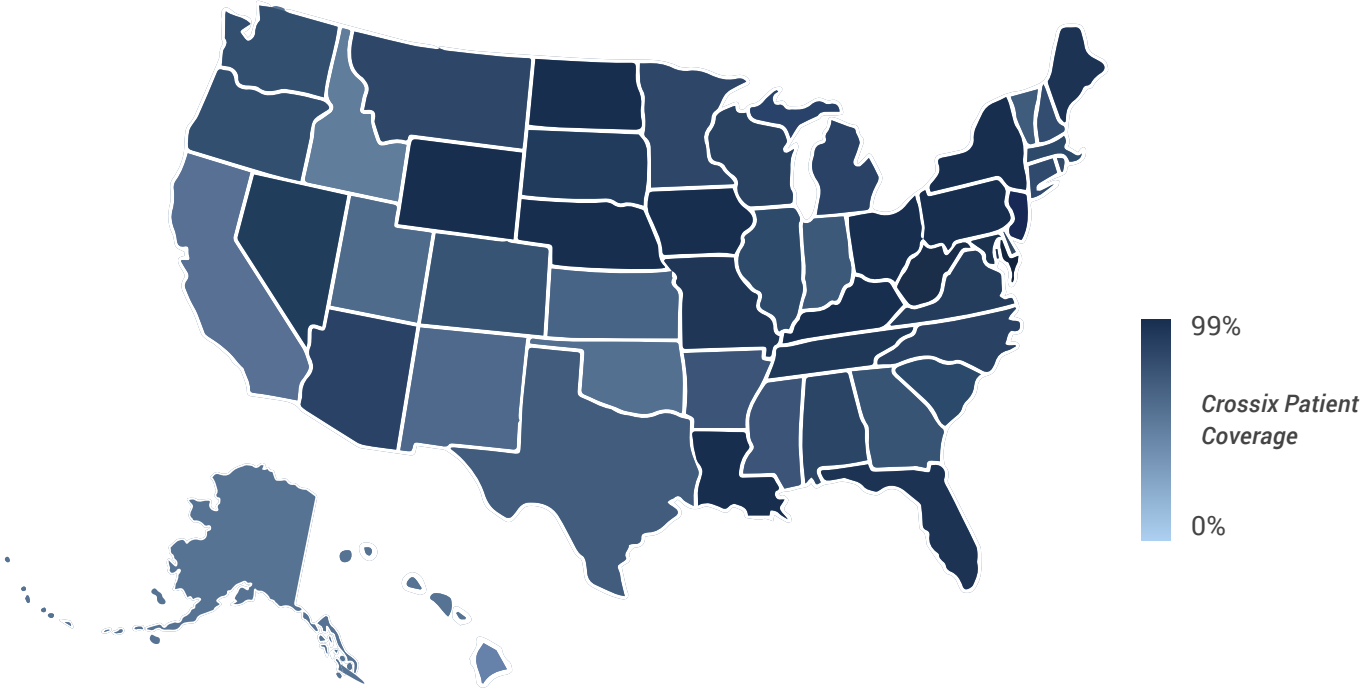
Coverage of US Health Data

Veeva Crossix has expansive coverage of US health data and provides complete transparency on coverage. The Crossix Data Platform covers 300MM+ patients – covering 75% of medical claims and 73% of prescription claims.

We provide specific counts for your area of interest upon request. Based on customer feedback, our counts are equal to or higher than counts available on other platforms more than 50% of the time. In head-to-head comparisons with other platforms, the Crossix Measurement Suite is selected 75% of the time, powered by the strong coverage of the Crossix Data Platform. The full set of prescription and medical claims in the Crossix Data Platform can be licensed to provide unlimited data access on demand using Veeva Compass Patient.

The Crossix Data Platform is built with modern, complex therapeutics in mind, leveraging multiple types of data including retail & specialty pharmacies, payors, and clearinghouses. Where others see challenges due to blocked products, Veeva Crossix's ability to view transactions through multiple lenses and better connect different health data sets together provides for superior visibility into these therapies.

CROSSIX DATA PLATFORM: 300MM+ PATIENT COVERAGE



Media Data

The Crossix Data Platform powers measurement and optimization for all major media investments, backed by a robust ecosystem of media integrations. We partner with leading media platforms and identity resolution providers to collect representative media exposure data across channels. We have also built the necessary match graph, operations and data infrastructure to request, ingest, normalize and analyze huge volumes of media owner first-party data.

MEASURED CHANNELS INCLUDE	
• TV (Set-top box data from multiple providers)	• Mail (Client & Third Parties)
• Digital (Cross Internet, Publishers)	• Savings Offers (Card Processors)
• Print (Subscriptions)	• Events (Client & Third Parties)
• POC (HCP Offices)	• EHR/HCP messaging (EHR Company)
• CRM (Client's DB Vendors)	• Sales Calls (Client)
• Email (Client & Third Parties)	

1st Party Connect

The 1st Party Connect program allows media owners to pass their first-party exposure data directly to Veeva Crossix or through a third-party processor, to be included in analytics. Crossix has built the necessary match graph, operations and data infrastructure to request, ingest, normalize and analyze huge volumes of media owner first-party data, which may vary widely across partners in terms of delivery method/cadence and file structure/contents. In general, Veeva Crossix requires media owners to pass timestamped, event-level data containing a matchable identity and campaign metadata.

Mapping and Grouping Media Data

Brands and agencies fully control how media data is mapped and grouped for measurement in the Crossix meaturement suite. Media plan and site mapping data is accessible on the platform.

Analytics Methodology

Analytics Methodology Overview

All Crossix measurement products use the same measurement methodology across every marketing channel; the only difference is how media exposure is collected. This provides a consistent measurement currency and accurate, scalable results across the enterprise.

Media exposure data is linked to health outcomes data by matching to the Crossix Data Platform. The intelligent matching process ensures that matching is accurate and that noise is not introduced via the matching process. We apply eligibility requirements to the matched sample to ensure a full longitudinal view of health data for every patient used in analysis (sample sizes are provided for every analysis).

Veeva Crossix measures incrementality through net impact analytics, which uses a Test vs. Control, observational study approach. We employ best in class methods for observational studies like those used for clinical trial analyses.

Metric Definitions

The Crossix Measurement Suite is highly configurable to answer specific business questions. Brand and agency stakeholders can align on and define key performance indicators (KPIs). The flexibility of the platform allows customizing KPIs to understand targets reached by media, how effectively they are reached, and what audience metrics may be predictive of conversion to brand.

All metric definitions are published on the platform. This includes the precise business logic underlying all metrics (for example, drugs or NDC codes, diagnosis codes, and time periods analyzed).

Feasibility

Feasibility analysis is a transparent process and standard requirement prior to any analysis.

Feasibility is an assessment that incorporates the Crossix Data Platform's counts for the brand and condition, benchmarks of media campaign performance, observed market rates, and media dynamics to assess confidence in Crossix's ability to deliver on client-specific KPIs at a particular moment in time.

When something is feasible, Crossix has confidence in the measurability of metrics for a specific campaign given the media volume being served.

Multiple inputs are considered during feasibility analysis including:

- Counts from Crossix Data Platform for brand and condition
- Media plan / impressions
- Campaign timing
- Natural market rate of relevant health events (e.g., diagnosis)
    - Smaller brands and conditions will require more media volume to be measured due to lower natural market incidence rates

Media Data Eligibility

Veeva Crossix applies eligibility criteria to media data to ensure accurate capture of exposed and unexposed audiences. Eligibility is channel specific. For example, for TV the eligibility requirement includes the following: households must view TV in each month of the campaign to be eligible for analysis. This excludes households that do not regularly turn on their TVs.

Defining Analysis Cohorts

Cohorts for analysis are defined based on media consumption metadata and the customer defined mapping and aggregation of media groups. We can aggregate and attribute results from these cohorts in a variety of ways, e.g., by ad type or tactic, to answer business questions. Our analytical techniques solve the problem of measuring publishers or channels entering later in the year, going dark, turning back on, or any of the other changes that occur in real-world execution of marketing campaigns. There is no need to further subdivide cohorts by exposure date.

Matching Exposures to Health Data

The Crossix Data Platform enables matches to be made using actual Personal Identifiable Information (PII) in a HIPAA-compliant way, leading to more accurate matches and results. This also means that all data matches and links are updated on the same timeframe as the data supplier – there is no lag to aggregate and process de-identified token data. The underlying health data that the analytics are run on are updated daily.

Because the Crossix Data Platform is able to make matches against actual PII in a privacy safe way, there is minimal risk of mis-identification or duplicate identification that comes with a tokenized approach. From an accuracy standpoint, we employ a probabilistic matching approach with many

factors involved. One example – our standard threshold for accuracy is 1/100 million. This means that if we take the combination of personal information available for matching (e.g., last name+first name+zipcode+DOB) and compare it to how frequently this occurs in the U.S. population, we will not match anything that is not at least unique to the degree of 1 in 100 million.

After media data is captured, it is sent throughout the distributed Veeva Crossix data network where it enters the data environments of the HIPAA covered entities that partner with Veeva Crossix. Each exposed individual then matches to either their own health records (DTC analyses) or to their patients' records (HCP analyses). The deidentified matched data results then return from each data node for processing and population into Crossix.

Health Data Eligibility

Veeva Crossix applies longitudinal eligibility criteria to all health data used in analysis in order to ensure sample robustness and representativeness.

We impose eligibility conditions to ensure sufficient longitudinal data coverage on every patient measured in the analysis. We take a conservative approach to ensure that we do not mistakenly attribute a conversion to a patient who only became visible to Veeva Crossix after exposure to marketing.

Longitudinal eligibility requirements for conversion metrics include:

- Rx in the year prior to exposure
- Rx within 6 months after exposure
- No Rx for the measured brand in the 6 months prior to exposure

Longitudinal eligibility requirements for audience metrics include:

- Rx in the year prior to exposure

Sample Sizes

Sample sizes and data waterfalls are available for all analyses. Additionally, robust granular feasibility is conducted before proceeding to any scope to ensure alignment on KPIs and business questions answered.

Measuring Incrementality

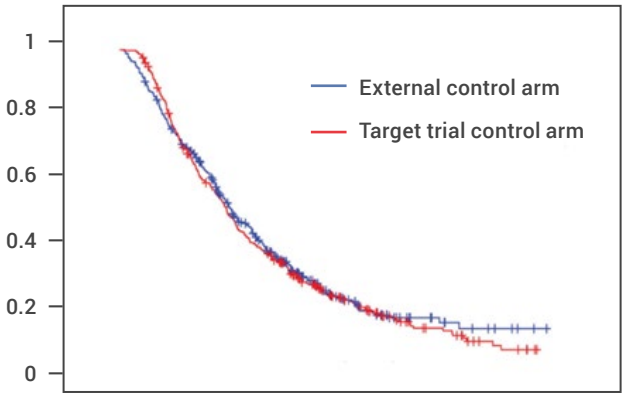
Crossix conducts test/control studies to isolate the impact of marketing and media interventions. By isolating the impact of these interventions, we help customers quantify the impact and ROI of marketing and media efforts, enabling brands to optimize accordingly.

The gold standard for any test/control study is a prospectively designed study with a randomized control. However, using a randomized control is not a common business practice due to cost and execution challenges. When a randomized control group is available, Crossix can use it for measurement. More commonly, Crossix conducts observational test/control studies to measure incrementality. This includes creating a control group and calculating the incremental impact of marketing as the difference between the behavior of the exposed (test) and control groups.

Control Group

Crossix uses propensity score matching to build control groups. This is a best practice, well studied methodology for accurately creating control groups for observational studies. It has been highlighted by the FDA as a methodology to inform clinical decisions based on real world evidence data.

Estimated survival over time of patients with non-small cell lung cancer in the target trial control arm and the external control arm created by propensity score matching.



Source: Designing Sound Clinical Trials That Incorporate Real-World Data, FDA.gov, 8/2022

The propensity score is a single metric that captures confounding variables. We calculate propensity scores using machine learning models that identify the confounders that need to be controlled for from a pool of more than 1,500 possible variables. We then match exposed individuals to control individuals based on the propensity score.

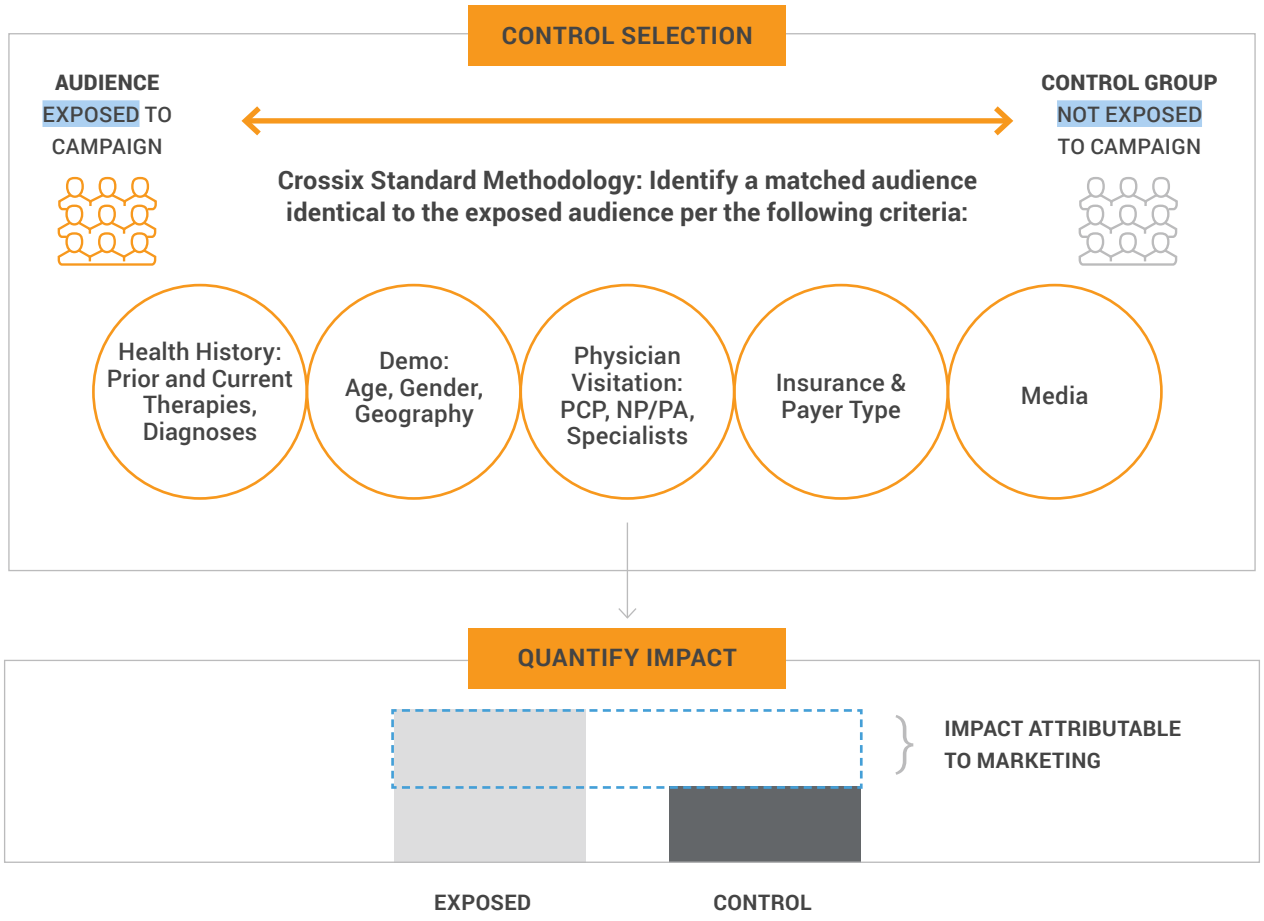
For control selection, Crossix establishes a 1:many matched pair control group (that is, for every individual exposed to media, multiple control individuals are chosen.) This is completed by identifying a matched audience identical to the exposed audience per criteria including treatment and diagnosis history, age, gender, geography, and media activity. The 1:many approach is more rigorous than a 1:1 approach since it results in larger sample sizes.

Veeva Crossix's approach is superior to other Test vs. Control approaches because of the rigor of the control group including eligibility rules, the number and variety of features controlled for, as well the unique capability to control for media activity within a given channel.

Compare Exposed and Control

We compare the post-exposure behavior of patients who saw media with the behavior of patients with similar health and media profiles who were not exposed to media. Since these two groups are equivalent with respect to their health propensities, we attribute the differences in behavior to the media exposure.

MATCHED PAIR CONTROLS ISOLATE IMPACT OF CAMPAIGN EXPOSURE



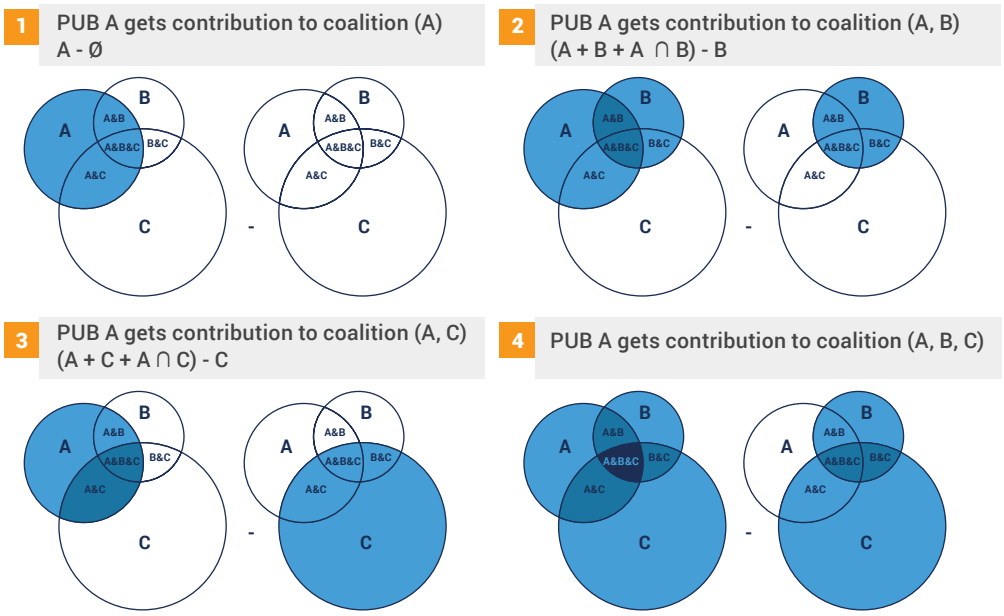
Attribution

Veeva Crossix employs best practice advanced analytics methods for attribution to enable accurately crediting media partners or channels for their impact on exposed patient behavior.

For attribution within a Net Impact analysis, Crossix uses the Shapley Value method. Crossix first analyzes each audience exposed to each channel, excluding overlap with other channels, to determine unique channel benefit. Then, each audience that overlaps between channels is analyzed to determine the overlap benefit, with results distributed based on the marginal contribution of the channel and timing (the gains of each marketing touchpoint). Lastly, the unique channel benefit is summed with the attributed overlap benefit to arrive at overall channel impact in addition to overall media impact. This approach is applied both within channels such as Digital and TV, as well as across channels and between DTC & HCP efforts.

Gross reporting uses universal attribution.

ATTRIBUTE VIA SHAPLEY VALUE METHOD



Example for a three publisher evaluation to calculate contribution of PUB A

1. We calculate PUB A's contribution (incremental conversions) to every possible publisher coalition (see chart above)
2. We sum these contributions, weight them by size, and take an average
3. We project this average up to each publisher's reach

Data Stream

Crossix Data Stream allows for all of the underlying data populated in Crossix to be pushed to a Crossix user for their own use in a flat file. Crossix Data Stream provides data at the same granularity available in Crossix along with several additional data sets. Crossix Data Stream can be shared with advertisers through SFTP or S3.

In addition to the data available in Crossix, Crossix Data Stream also includes:

- Zip3 DTC digital exposure data
- HCP NPI Feeds for eligible media partners
- Zip5 HCP digital exposure data
- Weekly TV Data

Conclusion

The Crossix Measurement Suite helps marketers maximize the impact of their media investments by connecting their campaigns to business outcomes using health data in a privacy-safe, HIPAA-compliant way. Transparency is built into the platform. Customers have a complete view of all inputs, business rules, and calculations that power Crossix. In this paper, we've reviewed these in depth with reviews of the Crossix Data Platform and our Analytics Methodology.

The Crossix Data Platform has robust coverage of US health data, covering 300MM+ patients, 75% of medical claims, and 73% of prescription claims. Critically, Crossix's health data is higher quality than what is available from other sources due to our better technology for linking data (SafeMine) and the unique composition of the data.

When it comes to health data that powers Crossix, Veeva shares everything – including the data itself. We provide detailed, granular counts for products or therapeutic areas upon request and as a standard part of feasibility assessment. The sample sizes for all analyses are provided. Furthermore, the full set of prescription and medical claims data that is used in Crossix is available for licensing through Veeva Compass and can be used for research and analysis that does not require the Crossix Measurement Suite.

The methodology that powers the Crossix Measurement Suite utilizes well studied and trusted techniques to generate actionable insights. Customers configure the applications to meet their specific needs – from how the marketing exposed population is cohorted for analysis to the definitions and business rules used for key performance indicators. These definitions are always available in Crossix for reference.

Together, the Crossix Data Platform and Crossix's Analytics Methodology are at the core of what has made the Crossix Measurement Suite the leading solution for measuring and optimizing healthcare marketing efforts.