Executive Summary

The rapidly expanding specialty pharmacy market has created unique challenges for pharmaceutical manufacturers looking to obtain dispensing, adherence, and switching data for specialty drugs. Access to timely, anonymized, longitudinal patient data forms the foundation of brand analytics within pharmaceutical manufacturers. While there are numerous syndicated sources for retail pharmacy data for traditional drugs, a comparable source for specialty drug data does not exist. Strong growth within the fragmented specialty pharmacy industry means that specialty drug data are now spread across a large number of pharmacies that include independent pharmacies, retailers/chains, hospitals, health plans, and pharmacy benefits managers (PBMs). In this landscape, the complete map of patient behavior is distributed across records residing in multiple specialty pharmacies. This whitepaper examines difficulties in merging specialty data from distinct specialty pharmacies, and offers a Health Insurance Portability and Accountability Act of 1996 (HIPAA)-compliant solution to de-identifying and merging specialty pharmacy data for manufacturer analysis.

Key Findings include:

- Specialty drugs have experienced impressive growth over the past decade; in 2007, specialty drugs accounted for 24% of net U.S. medicines spending, by 2016, specialty spending share hit 43%. Specialty drugs are distributed via specialty pharmacies, which in addition to dispensing specialty drugs, act as coordinated service organizations that manage the handling and support requirements unique to specialty drugs (e.g., patient education, adherence tracking, reimbursement support, etc.)

- The specialty pharmacy industry has experienced rapid growth, with one estimate indicating industry sales grew from $20 billion in 2005 to $115 billion in 2016. The rapid expansion in specialty pharmacy distribution decreases the likelihood that a single specialty pharmacy chain will have a complete record of a patient’s treatment history, making tracking, assessing, and monitoring specialty drug use more difficult.

- In lieu of a readily available syndicated source for specialty pharmacy data, manufacturers must proactively collect specialty pharmacy data from all of the specialty pharmacy locations that dispense their drug. However, merging separate specialty pharmacy data sets is hindered by patient privacy regulations (e.g., HIPAA), which require de-identification before sharing for non-clinical use.
The solution to HIPAA-compliant patient de-identification and data merging is through the use of unique, encrypted patient tokens, which allow the accurate matching of patient records across specialty pharmacies without exposure of the underlying PHI. Token-based matching supports the creation of a longitudinal record of a patient’s specialty prescriptions, regardless of where the drug was dispensed.

Specialty Drugs Replace Traditional Drugs as Growth Drivers

Specialty drugs have emerged as a key driver of pharmaceutical market growth. Over the past decade, specialty drug spending has nearly doubled, growing from 24% of net U.S. medicines spending in 2007 to 43% in 2016.\(^1\) Between 2015 and 2016 alone, specialty drugs drove 60.8% of new brand spending growth.\(^1\) Similarly, commercial customers of the pharmacy benefits manager (PBM), Express Scripts, saw a 13.3% increase in specialty drug spending between 2015 and 2016, compared with a 1.0% decrease in traditional medicines spending.\(^2\) Specialty market growth has been driven by a sharp rise in the number of newly available specialty medicines. Since 2010, specialty drugs have accounted for over half of the annual FDA new drug approvals (Figure 1).\(^3\) The recent boom in specialty drug development, commercialization, and utilization has garnered the interest of stakeholders across the pharmaceutical industry as key players attempt to define best practices in the still evolving specialty drug market.

**Figure 1. Annual Specialty New Drug Approvals Surpass Traditional Drug Approvals**

Sources: Based on Data from the PwC Health Research Institute (2007-2014); FDA New NME and BLA Approvals by Year; Express Scripts, CVS Specialty, Cigna, McKesson, and general web searches (2015 -2016).
Specialty Drugs are Expensive, High-Touch Drugs Used to Treat Rare or Complex Diseases

While there are no established criteria for what makes a drug a ‘specialty’ medicine, most stakeholders look for a combination of drug characteristics. Factors may include the drug’s clinical use to treat chronic, rare, or complex diseases (e.g., multiple sclerosis, Hepatitis C, oncology, etc.); a requirement for special handling or administration (e.g., injection, infusion, cold storage, etc.), availability via exclusive, restricted, or limited distribution; special patient monitoring for efficacy or safety; or high cost.\(^4,5\) Stakeholders agree that cost is an important factor, and for some, the only factor that determines specialty designation. For example, the Centers for Medicare & Medicaid Services (CMS) uses $670 per month as the dollar threshold for a specialty designation in Medicare Part D benefit.\(^5\)

Commercial spending data from Express Scripts highlights the striking difference in average costs between traditional medicines and specialty medicines. Based on 2016 data, the average cost of a prescription for a specialty therapy ranged from $1,500 to fill an HIV prescription to nearly $16,000 for a Hepatitis C prescription (Figure 2a).\(^6\) In stark contrast, the most expensive traditional therapy area in 2016 was in attention disorders, where an average prescription cost only $145.

Specialty drugs are typically more complex than traditional small molecule drugs, and while there is a trend toward more specialty oral agents, the majority are formulated as an injection or infusion. To re-coup the high costs associated with specialty drug development and commercialization, manufacturers price specialty agents much higher than traditional drugs. One subclass of specialty drugs includes the biologics, which are typically large, complex molecules or mixtures of molecules produced from a living system (e.g., bacterial, plant or animal cells). Biologics’ manufacturing complexity, as opposed to a chemically-synthesized small molecule drug, have made these agents largely resistant to the downward pricing pressure from competing biosimilars.
Figure 2. Express Scripts 2016 Commercial Utilization and Unit Costs Across Top Ten Therapy Classes

<table>
<thead>
<tr>
<th>Specialty Therapy Class</th>
<th>Average Commercial Rx Price</th>
<th>Prevalance of Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hepatitis C</td>
<td>$15,708.27</td>
<td>0.03%</td>
</tr>
<tr>
<td>Oncology</td>
<td>$7,890.81</td>
<td>0.1%</td>
</tr>
<tr>
<td>Multiple Sclerosis</td>
<td>$5,055.80</td>
<td>0.1%</td>
</tr>
<tr>
<td>Inflammatory Conditions</td>
<td>$3,587.83</td>
<td>0.4%</td>
</tr>
<tr>
<td>HIV</td>
<td>$1,555.56</td>
<td>0.2%</td>
</tr>
<tr>
<td>Attention Disorders</td>
<td>$145.45</td>
<td>2.9%</td>
</tr>
<tr>
<td>Skin Conditions</td>
<td>$145.21</td>
<td>7.0%</td>
</tr>
<tr>
<td>Diabetes</td>
<td>$125.82</td>
<td>5.3%</td>
</tr>
<tr>
<td>Asthma</td>
<td>$68.86</td>
<td>8.9%</td>
</tr>
<tr>
<td>Pain/inflammation</td>
<td>$48.85</td>
<td>22.0%</td>
</tr>
</tbody>
</table>

Source: Based on data from Express Scripts 2016 Drug Trend Report.

The difference in prescription costs between traditional and specialty medicines also relates to the size of the drug’s target population. Specialty drugs often treat rare or complex diseases, which affect a much smaller proportion of the population when compared with traditional small-molecule drugs that treat more common diseases like hypertension or high cholesterol. As shown in Figure 2b, Hepatitis C had the most expensive average prescription cost ($15,708.27) and the lowest prevalence rate (0.03%) among Express Scripts commercial customers. Of the five specialty therapy classes included, all five had a population prevalence of less than 0.5%. In contrast, all of the traditional therapy classes had an average prevalence that ranged from 2.9% to 22% of the commercial population. This small treatable population is a key reason for the high price of specialty agents, as the cost of development must be spread over a much smaller base of paying customers than for common diseases.

**With Low Disease Prevalence and High Drug Costs, Specialty Patients Are High Value**

Given the smaller target populations for specialty drugs, from the perspective of the drug manufacturer,
the dollar value of a patient taking a specialty drug is higher than that of a patient taking a traditional small molecule drug. Therefore, it is critical for manufacturers to understand when patients are non-adherent to a specialty medication, or when they make a therapy change to a competing product.

However, traditional drug volume and prescription data (i.e., retail pharmacy data), despite being widely available, is not ideal for tracking specialty drugs. Most specialty drugs require additional pharmacy services, support, or reporting that are outside of the capabilities of retail pharmacies. As a result, specialty drugs are instead distributed through a smaller, more fragmented network of specialty pharmacies that have the capabilities needed to support specialty drug patients.

**Specialty Pharmacies: A Rapidly Expanding Industry**

The National Association of Specialty Pharmacy (NASP) defines a specialty pharmacy as “a state-licensed pharmacy that solely or largely provides only medications for people with serious health conditions requiring complex therapies.” In addition to its primary role as a distributor of specialty pharmaceuticals, specialty pharmacies are coordinated service organizations created to manage the handling and service requirements of specialty drugs. The services are patient-centric, high-touch, often individualized and may include elements of patient education, drug delivery training, patient monitoring, caregiver communication support, case management, and reimbursement support. Some specialty pharmacies also manage insurer reimbursement paperwork, manufacturer data reporting, and requirements for FDA data reporting. The end goal of the specialty pharmacy model is to improve access to care, clinical outcomes, and economic outcomes.

The specialty pharmacy industry has experienced rapid growth, with estimates by Pembroke Consulting revealing industry sales grew from $20 billion in 2005 to $115 billion in 2016. The industry’s strong growth has attracted a range of specialty pharmacy players, including independent pharmacies, retailers/chains, hospitals, health plans, and PBMs all looking to start, acquire, or expand specialty pharmacy capabilities. In 2008, the Utilization Review Accreditation Commission (URAC), a leading independent accreditation organization, listed only a handful of fully accredited specialty pharmacies; today, there are 552 fully accredited specialty pharmacies listed in the URAC database.

**A Fractured Specialty Landscape Presents Data Acquisition Challenges for Specialty Manufacturers**

As a result of the rapid industry growth, the specialty pharmacy landscape has become fractured with many new players. While the top ten specialty pharmacies accounted for three-quarters of specialty prescription revenues in 2016 (Figure 3), the remaining one-quarter of specialty prescription sales may be spread across over 500 specialty pharmacies.
One example in which the complexity of specialty drug distribution and management creates considerable challenges is for drug manufacturers who want to collect and analyze patient utilization data. The rapid expansion in specialty pharmacy locations decreases the likelihood that a single specialty pharmacy will have a complete record of a patient’s treatment history, making tracking, assessing, and monitoring specialty drug use more difficult. Likewise, no third-party aggregator(s) exist to gather this information proactively to make it available for resale to manufacturers as exists in the retail pharmacy space.

Rather than having a single source of timely utilization data, specialty patient data must be gathered from all of the specialty pharmacy locations that dispense a specialty medication. This aggregation places the onus squarely on the manufacturer to merge the disparate data sets, and ensure that the data are complete, properly anonymized, and formatted for integration with internal data platforms.

### Compliance Considerations for Manufacturers Working with Specialty Pharmacy Data

Merging data from separate specialty pharmacies is complicated by the need for healthcare organizations...
to comply with privacy regulations outlined in the Health Insurance Portability and Accountability Act of 1996 (HIPAA), and the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009, which outline the federal standards for privacy and security of protected health information (PHI). Specialty pharmacy data are full of PHI elements (e.g., names, addresses, dates of birth, etc.). To ensure that manufacturers avoid a HIPAA violation, PHI must be removed through a process of de-identification before merging separate healthcare data sets that are intended for non-clinical use (e.g., for brand analysis, tracking patient journeys, identifying switch patterns, etc.).

Removing PHI data elements from specialty data sets presents a unique challenge for manufacturers who aim to create a single, merged set of specialty dispensing data. To avoid non-compliant exposure to PHI, manufacturers would need to receive de-identified patient data from each specialty pharmacy that distributes their drug. However, once a data set is de-identified, the manufacturer would have no way to link a patient’s records across specialty pharmacy data sets.

For example, one metric that manufacturers track is the patient discontinuation rate. Datasets that have been merged after they were de-identified may unknowingly contain partial records that originated from two separate specialty pharmacies for the same patient. (Figure 4). If a patient stops filling a prescription at one specialty pharmacy, an analyst using purely de-identified data would be unable to determine whether that patient discontinued the drug, or whether they simply switched to a different specialty pharmacy to fill their next prescription. In this example, the patient discontinuation rate would appear artificially high. For a manufacturer to obtain an accurate snapshot of longitudinal patient records, and to effectively triage patients for service interventions, they need to accurately join data from all of the specialty pharmacies that distribute their specialty drug, and track unique, but de-identified, patients as they move within the network.
HIPAA-Compliant De-Identification and Linking of Specialty Pharmacy Data Using Patient Tokens

The solution to HIPAA-compliant patient de-identification and data merging is through the use of encrypted patient tokens that are reliably produced no matter where de-identification occurs, a method that allows the accurate matching of patient records across specialty pharmacies without exposure of the underlying PHI. The use of unique patient tokens allows a manufacturer’s analysts to create a longitudinal record of each patient’s prescription record, regardless of where the specialty drug is dispensed.

Datavant Technology for De-Identification, Tokenization, and Data Linking

Datavant is an ideal solution for compliant de-identification and data linking. The Datavant de-identification engine processes raw specialty pharmacy data, removing all elements of PHI (e.g., names, dates of birth, etc.), and creates a statistically de-identified data file that meets HIPAA requirements under the expert determination method.

During the de-identification process, Datavant also adds one or more unique encrypted patient tokens to each record. These tokens are irreversible 44-character long strings that replace the patient’s PHI and serves as a unique ID for the person, and is reliably reproduced for the same patient regardless of where the software is run.

For clients for whom Datavant supports the merging of specialty pharmacy data, this process of data de-identification, tokenization, and linking can occur through two data flows (Figure 5), both of which result in the creation of an accurately linked, de-identified (PHI free), and tokenized data set. The primary difference between the data flows is the location at which the Datavant technology is run (Datavant technology can be installed and run at any site: specialty pharmacies, EMRs, claims clearinghouses, trusted third parties (TTP), etc.).

A The first data flow is used by data sources who don’t want to send PHI outside of their environment. In these cases, the specialty pharmacy installs Datavant on site to de-identify and tokenize their patient data, and then sends out the de-identified data with encrypted patient tokens either directly to the manufacturer or to a TTP for aggregation.

B The second data flow is used by data sources who don’t mind sending out PHI under a...
Business Associate Agreement (BAA). In these cases, the specialty pharmacy sends the raw patient data (with PHI) to a TTP with a HIPAA-compliant data center. Using Datavant within this data center, the TTP then de-identifies and tokenizes the data to prepare it for aggregation.

Once all of the specialty data have been de-identified and tokenized using the Datavant de-identification engine, the TTP can find and link patient records across the different specialty data sets. Importantly, Datavant’s method of data linking does not expose the manufacturer to PHI: the unique encrypted tokens, not the PHI, are what are used to find and link patients across disparate data sets. Once complete, the manufacturer’s analysts receive a single merged specialty data set, and can be certain that the records for each unique patient are both PHI-free and longitudinally complete, regardless of where the patient filled a prescription. This process of creating unique patient tokens allows a manufacturer to overcome two important hurdles: avoiding PHI exposure (and remaining HIPAA-compliant), and achieving accurate data linking across disparate specialty data sets.

The Benefits of Merged Specialty Pharmacy Data

The recent rise in both the number and the type of specialty pharmacies has made it more difficult for specialty drug manufacturers to accurately track their specialty drug patients. The use of encrypted, unique patient tokens provides manufacturers a HIPAA-compliant method of joining patient records that
are split across multiple specialty pharmacy locations, without bringing PHI into the data environment (Figure 6).

**Figure 6. Longitudinal, De-Identified Specialty Pharmacy Data Sets for Accurate Patient Targeting**

Datavant technology for de-identifying and linking patient records provides manufacturers a tool to create a single, joined specialty pharmacy data set. With access to the complete data story, specialty drug manufacturers can quickly, easily, and accurately triage their specialty drug patients for the appropriate intervention. For manufacturers, the increased visibility into how patients are using their drug means better service and support for new patients, for non-adherent or discontinuing patients, for patients experiencing side-effects, and for patients facing reimbursement challenges.

With a single, merged dataset, manufacturers can confidently track:

- **New Patient Starts**: Tracking will identify true ‘new starts’, and not inaccurately target patients who had been receiving prescriptions previously at a different pharmacy network.
- **Adherence Measurements**: Lagging or missed refills can be flagged for a service intervention, without the concern that a patient is simply switching to a different pharmacy network.
- **Discontinuation and Patient Switching**: Only patients who are truly falling off therapy or switching to a new therapy are targeted for intervention.

In each of these tracking scenarios, an analyst who spots a treatment issue can send the unique patient token back to the specialty pharmacy, who can map the token back to the original patient, and then initiate a service intervention (Figure 7).
Figure 7. Longitudinal, De-Identified Specialty Pharmacy Data Sets for Accurate Patient Targeting

This improved access to linked specialty pharmacy data allows manufacturers to retain market share by helping patients find ways to stay on therapy, and ensures that patients are receiving the support they need to achieve the best health outcomes possible.

Sources

For more information:

- Contact Jason LaBonte, Ph.D. for questions or comments about this analysis:
  Jason@datavant.com

- Contact Lauren Stahl for more information about the Datavant modules that were used in this study:
  Lauren@datavant.com

- Visit the Datavant website to read our other whitepapers and materials:
  www.datavant.com

Organizing the World’s Health Data

Datavant helps organizations safely share and link healthcare data.

We believe in connecting healthcare data to eliminate the silos of healthcare information that hold back innovative medical research and improved patient care. We help data owners manage the privacy, security, compliance, and trust required to enable safe data sharing.

Datavant’s vision is backed by Roivant Sciences, Softbank, and Founders Fund, and combines technical leadership and healthcare expertise. Datavant is located in the heart of San Francisco’s Financial District.
Glossary of Terms:

Covered Entity
A covered entity (CE) under HIPAA is a health care provider (e.g. doctors, dentists, pharmacies, etc), a health plan (e.g. private insurance, government programs like Medicare, etc), or a health care clearinghouse (i.e. entities that process and transmit healthcare information).

De-identified health data
De-identified health data is data that has had PII removed. Per the HIPAA Privacy Rule, healthcare data not in use for clinical support must have all information that can identify a patient removed before use. This rule offers two paths to compliantly remove this information: the Safe Harbor method and the Statistical method. When these identifying elements have been removed, the resulting de-identified health data set can be used without restriction or disclosure.

Deterministic matching
Deterministic matching is when fields in two data sets are matched using a unique value. In practice, this value can be a social security number, Medicare Beneficiary ID, or any other value that is known to only correspond to a single entity. Deterministic matching has higher accuracy rates than probabilistic matching, but is not perfect due to data entry errors (mis-typing a social security number such that matching on that field actually matches two different individuals).

Encrypted patient token
Encrypted patient tokens are non-reversible 44 character strings created from a patient’s PHI, allowing a patient’s records to be matched across different de-identified health data sets without exposure of the original PHI.

False positive
A false positive is a result that incorrectly states that a test condition is positive. In the case of matching patient records between data sets, a false positive is the condition where a “match” of two records does not actually represent records for the same patient. False positives are more common in probabilistic matching than in deterministic matching.

Fuzzy matching
Fuzzy matching is the process of finding values that match approximately rather than exactly. In the case of matching PHI, fuzzy matching can include matching on different variants of a name (Jamie, Jim, and Jimmy all being allowed as a match for “James”). To facilitate fuzzy matching, algorithms like SOUNDEX can allow for differently spelled character strings to generate the same output value.

Health Information Technology for Economic and Clinical Health (HITECH) Act
The HITECH Act was passed as part of the as part of the American Recovery and Reinvestment Act of 2009 (ARRA) economic stimulus bill. HITECH was designed to accelerate the adoption of electronic medical records (EMR) through the use of financial incentives for “meaningful use” of EMRs until 2015.
and financial penalties for failure to do so thereafter. HITECH added important security regulations and data breach liability rules that built on the rules laid out in HIPAA.

Health Insurance Portability and Accountability Act of 1996 (HIPAA)
HIPAA is a U.S. law requiring the U.S. Department of Health and Human Services (HHS) to develop security and privacy regulations for protected health information. Prior to HIPAA, no such standards existed in the industry. HHS created the HIPAA Privacy Rule and HIPAA Security Rule to fulfill their obligation, and the Office for Civil Rights (OCR) within HHS has the responsibility of enforcing these rules.

Personally-identifiable information (PII)
Personally-identifiable information (PII) is a general term in information and security laws describing any information that allows an individual to be identified either directly or indirectly. PII is a U.S.-centric abbreviation, but is generally equivalent to “personal information” and similar terms outside the United States. PII can consist as informational elements like name, address, social security number or other identifying number or code, telephone number, email address, etc., but can include non-specific data elements such as gender, race, birth date, geographic indicator, etc. that together can still allow indirect identification of an individual.

Probabilistic matching
Probabilistic matching is when fields in two data sets are matched using values that are known not to be unique, but the combination of values gives a high probability that the correct entity is matched. In practice, names, birth dates, and other identifying but non-unique values can be used (often in combination) to facilitate probabilistic matching.

Protected health information (PHI)
Protected health information (PHI) refers to information that includes health status, health care (physician visits, prescriptions, procedures, etc.), or payment for that care and can be linked to an individual. Under U.S. law, PHI is information that is specifically created or collected by a covered entity.

Safe Harbor de-identification
HIPAA guidelines requiring the removal of identifying information offered covered entities a simple, compliant path to satisfying the HIPAA Privacy Rule through the Safe Harbor method. The Safe Harbor de-identification method is to remove any data element that falls within 18 different categories of information, including:

1. Names
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes. However, you do not have to remove the first three digits of the ZIP code if there are more than 20,000 people living in that ZIP code.
3. The day and month of dates that are directly related to an individual, including birth date, date of admission and discharge, and date of death. If the patient is over age 89, you must also remove his age and the year of his birth date.
4. Telephone number  
5. Fax number  
6. Email addresses  
7. Social Security number  
8. Medical record number  
9. Health plan beneficiary number  
10. Account number  
11. Certificate or license number  
12. Vehicle identifiers and serial numbers, including license plate numbers  
13. Device identifiers and serial numbers  
14. Web addresses (URLs)  
15. Internet Protocol (IP) addresses  
16. Biometric identifiers, such as fingerprints  
17. Full-face photographs or comparable images  
18. Any other unique identifying number, such as a clinical trial number

Social Security Death Master File
The U.S. Social Security Administration maintains a file of over 86 million records of deaths collected from social security payments, but it is not a complete compilation of deaths in the United States. In recent years, multiple states have opted out of contributing their information to the Death Master File and its level of completeness has declined substantially. This Death Master File has limited access, and users must be certified to receive it. This file contains PHI elements like social security numbers, names, and dates of birth — therefore, bringing the raw data into a healthcare data environment could risk a HIPAA violation.

Soundex
Soundex is a phonetic algorithm that codes similarly sounding names (in English) as a consistent value. Soundex is commonly used when matching surnames across data sets as variations in spelling are common in data entry. Each soundex code generated from an input text string has 4 characters – the first letter of the name, and then 3 digits generated from the remaining characters, with similar-sounding phonetic elements coded the same (e.g. D and T are both coded as a 3, M and N are both coded as a 5).

Statistical de-identification (also known as Expert Determination)
Because the HIPAA Safe Harbor de-identification method removes all identifying elements, the resulting de-identified health data set is often stripped of substantial analytical value. Therefore, statistical de-identification is used instead (HIPAA calls this pathway to compliance “Expert Determination”). In this method, a statistician or HIPAA certification professional certifies that enough identifying data elements have been removed from the health data set that there is a “very small risk” that a recipient could identify an individual. Statistical de-identification often allows dates of service to remain in de-identified data sets, which are critical for the analysis of a patient’s journey, for determining an episode of care, and other common healthcare investigations.