JPMSA
JOURNAL OF THE PHARMACEUTICAL MANAGEMENT SCIENCE ASSOCIATION

SPRING 2025 | IN THIS ISSUE:

pmsa
PHARMACEUTICAL MANAGEMENT SCIENCE ASSOCIATION

Infusing smarts and heart in every project. It matters.

HEART MATTERS.

At ZS, we're committed to helping our clients improve life and how we live it. As a global consulting and technology firm, ZS infuses our deep expertise and human ingenuity to help our clients drive digital transformation, apply AI to real-world problems and keep innovation flowing. Together, let's make a heartfelt impact where it matters.

ZS.COM/ITMATTERS
Impact where it matters.®

ZS

Table of Contents

Official Publication of the Pharmaceutical Management Science Association (PMSA)

The mission of the Pharmaceutical Management Science Association not-for-profit organization is to efficiently meet society's pharmaceutical needs through the use of management science.

The key points in achieving this mission are:

iii. Raise awareness and promote use of Management Science in the pharmaceutical industry

iii. Foster sharing of ideas, challenges, and learning to increase overall level of knowledge and skill in this area

iii. Provide training opportunities to ensure continual growth within Pharmaceutical Management Science

iii. Encourage interaction and networking among peers in this area

Please submit correspondence to:
Pharmaceutical Management Science Association
1024 Capital Center Drive, Suite 205
Frankfort, KY 40601
info@pmsa.org
(877) 279-3422

Executive Director:
Krystal Livingston
klivingston@pmsa.org

PMSA Board of Directors

The *Journal of the Pharmaceutical Management Science Association (ISSN 2473-9685 (print) / ISSN 2473-9693(online))* is published annually by the Pharmaceutical Management Science Association.

PMSA Journal: Spotlighting Analytics Research

PMSA is pleased to announce the 2025 Journal of the Pharmaceutical Management Science Association (PMSA), the official research publication of PMSA

The Journal publishes manuscripts that advance knowledge across a wide range of practical issues in the application of analytic techniques to solve Pharmaceutical Management Science problems, and that support the professional growth of PMSA members. Articles cover a wide range of peer-reviewed practice papers, research articles and professional briefings written by industry experts and academics. Articles focus on issues of key importance to pharmaceutical management science practitioners.

If you are interested in submitting content for future issues of the Journal, please send your submissions to info@pmsa.org.

GUIDELINES FOR AUTHORS
Summary of manuscript structure: An abstract should be included, comprising approximately 150 words. Six key words are also required. All articles and papers should be accompanied by a short description of the author(s) (approx. 100 words).

Industry submissions: For practitioners working in the pharmaceutical industry, and the consultants and other supporting professionals working with them, the Journal offers the opportunity to publish leading-edge thinking to a targeted and relevant audience.

Industry submissions should represent the work of the practical application of management science methods or techniques to solving a specific pharmaceutical marketing analytic problem. Preference will be given to papers presenting original data (qualitative or quantitative), case studies and examples. Submissions that are overtly promotional are discouraged and will not be accepted.

Industry submissions should aim for a length of 3000-5000 words and should be written in a 3rd person, objective style. They should be referenced to reflect the prior work on which the paper is based. References should be presented in Vancouver format.

Academic submissions: For academics studying the domains of management science in the pharmaceutical industry, the Journal offers an opportunity for early publication of research that is unlikely to conflict with later publication in higher-rated academic journals.

Academic submissions should represent original empirical research or critical reviews of prior work that are relevant to the pharmaceutical management science industry. Academic papers are expected to balance theoretical foundations and rigor with relevance to a non-academic readership. Submissions that are not original or that are not relevant to the industry are discouraged and will not be accepted.

Academic submissions should aim for a length of 3000-5000 words and should be written in a third person, objective style. They should be referenced to reflect the prior work on which the paper is based. References should be presented in Vancouver format.

Expert Opinion Submissions: For experts working in the Pharmaceutical Management Science area, the Journal offers the opportunity to publish expert opinions to a relevant audience.

Expert opinion submissions should represent original thinking in the areas of marketing and strategic management as it relates to the pharmaceutical industry. Expert opinions could constitute a review of different methods or data sources, or a discussion of relevant advances in the industry.

Expert opinion submissions should aim for a length of 2000-3000 words and should be written in a third person, objective style. While references are not essential for expert opinion submissions, they are encouraged and should be presented in Vancouver format.

Industry, academic and expert opinion authors are invited to contact the editor directly if they wish to clarify the relevance of their submission to the Journal or seek guidance regarding content before submission. In addition, academic or industry authors who wish to cooperate with other authors are welcome to contact the editor who may be able to facilitate useful introductions.

Thank you to the following reviewers for their assistance with this issue of the *PMSA Journal:*

Pushpendra Arora, Merck
Kanchana Chandra , Amgen
Nathan Corder, Eli Lilly and Company
Hetu Gadhia, Merck
Ewa Kleczyk, Target RWE
Wendy Lawhead, Pfizer
Eren Sakinc, Bayer
J.P. Tsang, Bayser
Ray Wolson, Sandoz
Devesh Verma, Axtria
James Young, UCB

Editor: Srihari Jaganathan, UCB

Using Longitudinal Claims Data to Predict Response Uptake in Rare and Ultra-Rare Diseases

Jerry A. Rosenblatt, PhD – Rosenblatt Life Science Consultants
Adrian Ghenadenik, PhD – Rosenblatt Life Science Consultants
Rhys Rosenberg – Rosenblatt Life Science Consultants
Ashwin Anand – FORIAN Inc.

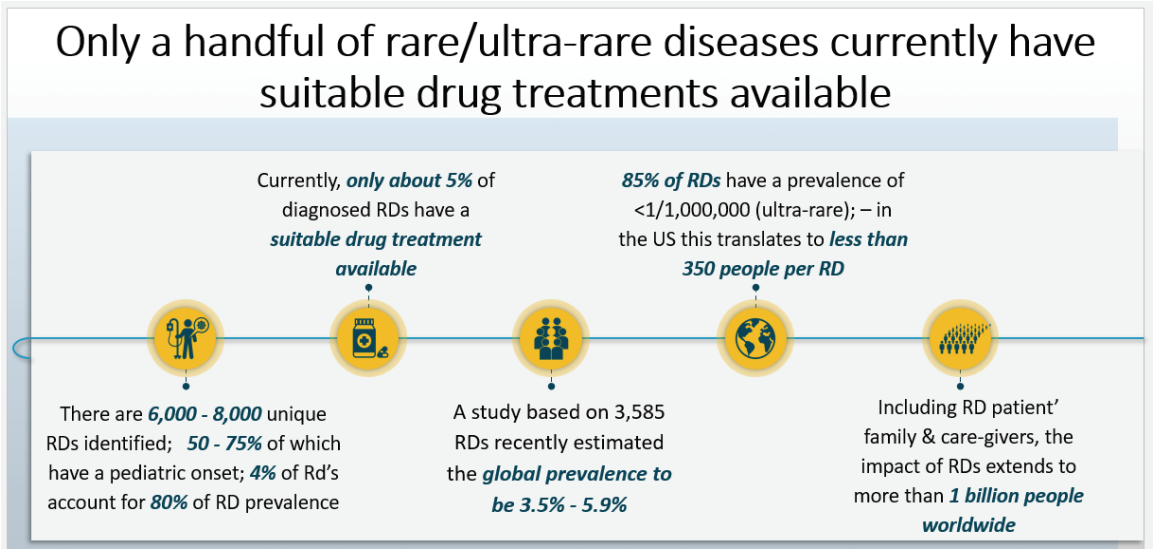
Abstract: The development of treatments for rare and ultra-rare diseases faces several key challenges. These include limited knowledge about the diseases themselves, the complexity involved in obtaining robust clinical evidence, and the inherent difficulties associated with recruiting sufficient numbers of patients for clinical trials. Moreover, the development of these treatments requires significant investment in research and development. The objectives of this study were to identify and define the potential attributes of both products and markets that impact the uptake of treatments in rare disease markets. This was achieved by analyzing longitudinal medical and patient claims data, allowing us to develop a better understanding of the market dynamics for these types of diseases.

Introduction

Rare diseases (RDs) are increasingly a priority area of focus for both public health and the pharmaceutical industry. There is no single definition of what constitutes a rare disease; in the United States, the most common threshold is based on the 1983 Orphan Drug Act,[1] which defines RDs as those having a prevalence of less than 200,000. Many jurisdictions, including the European Union and Japan, use definitions somewhat similar to that of the US, typically fluctuating around 4-5/10,000 (vs. 6.4/10,000 in the US), nonetheless, other thresholds differ significantly, for example in Australia, where a 1.2/10,000 limit is used.[2]

The number of RDs is not exactly known, and estimates vary widely. The reasons behind this variability include the lack of consistency in defining prevalence limits and discrete disease entities, significant differences in incidence among countries and jurisdictions, and inconsistencies in terminology related to what constitutes a rare disease.[3,4] A 2020 analysis of the OrphaNet database identified 6,172 clinically unique RDs,[5] within a similar range to the 5,000-8,000 cited in other sources.[3,6] Approximately 4% of RDs account for about 80% of the cumulative RD disease prevalence, while the remainder is largely composed of extremely infrequent conditions: about 85% of RDs have a point prevalence of <1/1,000,000,[5] and therefore may be categorized as ultra-rare.

Figure 1: Characteristics of rare disease



The majority of RDs are thought to be primarily of genetic origin, and between 50% to 75% of them have a pediatric onset.[7] Furthermore, while the prevalence of any given RD is very limited, estimates suggest a cumulative global prevalence ranging between 3.5% to 5.9%; actual prevalence may be even higher.[3,7] When considering patients, their families, and caregivers, rare diseases are estimated to impact more than 1 billion people worldwide. The burden of RDs is high: studies conducted in the past two decades estimate that between 20% and 60% of deaths among newborns, infants, and children are due to rare diseases. Furthermore, approximately 55% to 65% of RDs are associated with significantly reduced life expectancy.[6]

Regardless of their individual characteristics, RDs are, at best, challenging to manage from a clinical perspective. Very few physicians have enough experience to diagnose and manage most RDs, and even when properly diagnosed, appropriate clinical management is difficult to achie[3]. Incorrect or delayed diagnoses occur frequently, typically requiring visits with numerous healthcare professionals, several diagnostic tests, and in many cases, going through several inadequate medical treatments.[8,9] A recent European survey covering 41 countries reported a 4.7-year median time to diagnosis, with 25% of patients experiencing even longer delays.[8]

Currently, only about 5% of rare diseases have approved pharmacological treatments,[10,11] highlighting the need for further research and development in this area. Nonetheless, due to an increasing focus on RDs, the number of orphan drug approvals has grown exponentially, from 93 treatments garnering initial approval between 1990 to 1999, to a similar number of drug approvals in only three years, between 2020 and 2022.[12] Moreover, more than half of all new drugs approved by the FDA between 2018 and 2023 had an orphan drug designation.[13] To date, almost 40% of orphan drug approvals have been for rare oncological diseases; however, more recently other therapeutic areas, such as cardiology, hematology, endocrinology, and nephrology have seen considerable growth in drug approvals.[10,12,13]

Research and development of treatments for rare diseases is particularly challenging due to several factors. First, the sheer number of RDs results in limited knowledge about any individual condition, including their natural history, diagnosis, and potential treatment pathways. Second, their small prevalence and population diversity often limits clinical trial design options and makes patient recruitment inherently difficult. Finally, given the reasons described above, developing treatments for RDs requires considerable investment, which is often associated with relatively low probability of success.

Several efforts are being deployed by the FDA to foster development of RD treatments. These include 1) CDER's (Center for Drug Evaluation and Research) Accelerating Rare disease Cures (ARC) program, seeking to provide strategic direction and coordination of CDER's activities in RDs, 2) the establishment of a Rare Disease Innovation Hub, focusing on conditions where the natural history of disease is not well understood, 3) the launch of a Genetic Metabolic Diseases Advisory Committee, with the objective of advising the FDA on the efficacy and safety of drugs targeting genetic metabolic diseases, 4) the implementation of the START pilot program, designed to provide knowledge seeking facilitate more efficient development of potentially life-saving RD treatments, and 5) the RDEA pilot program, designed to support the development of efficacy endpoints.[13,14]

Research Objectives

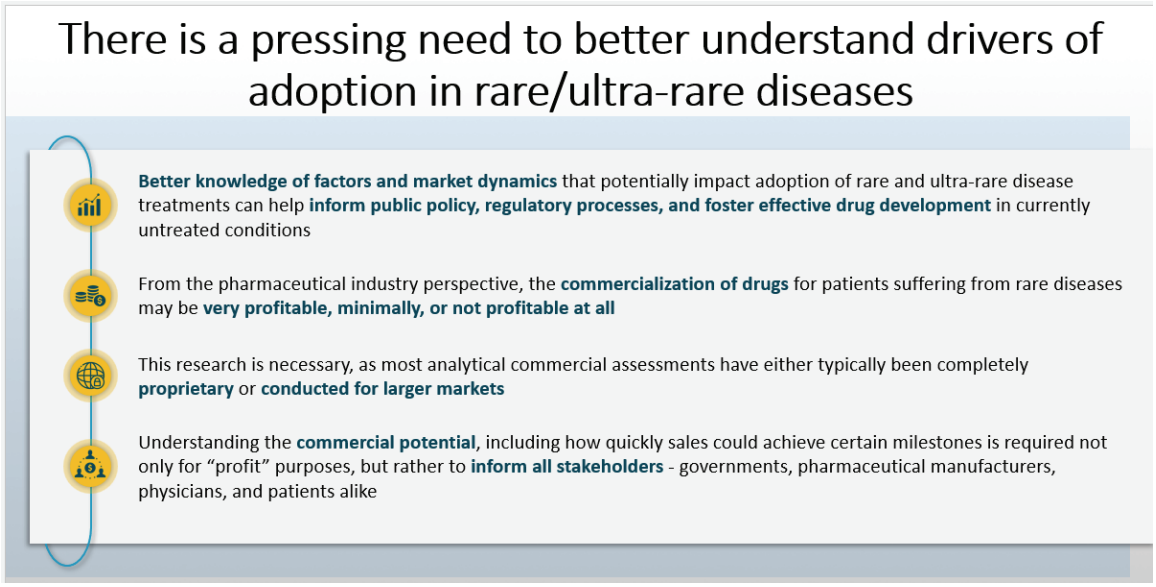


Figure 2: Objectives of analysis

The complex and poorly understood characteristics of rare diseases make research in this area a priority. Furthermore, given the inherent risks associated with drug development for RDs highlight the need to better understand the factors that may influence the uptake of drugs in this space. Therefore, the objective of this study was to identify

and define the potential attributes of both products and markets that impact the uptake of treatments in rare disease markets. Specifically, using longitudinal medical and patient claims data, we sought to identify potential drivers of product adoption, focusing on product and market attributes that influence the shape and timing of uptake curves. In addition, we aimed to develop insights into how different market dynamics, such as company size, product efficacy, and unmet medical needs, affect the speed of adoption.

Methodology

Product Selection

This study utilized a HIPAA-compliant anonymized, patient-level longitudinal medical and pharmaceutical hybrid claims database from Forian Inc.

Establishing the list of products to be considered in this analysis required a qualitative assessment of products launched in the past 10 years, evaluating whether the product fit the criteria of a launch in the rare disease space. Using the FDA rare disease definition, utilization and patient claim data for all products in the Forian CHRONOS database was examined. Since data gaps may exist in patients claims data, and sample sizes may be too small in some indications, products included in these analyses were selected based on the completeness of data records and sufficient number of patient records. In addition, only products that had at least 2 years of historical claims were included. The final selection included 38 products.

Qualitative and Quantitative Product Assessment

Each product was characterized using a list of product attributes based on their likelihood to impact product uptake. Clinical, commercial and disease-level variables were evaluated using a combination of qualitative and quantitative research to develop product profiles for the various prospective launches of interest.

Patient Curve Analysis

Using the Forian CHRONOS patient claims database, patient treatment journeys were constructed using longitudinal identifiers at the pharmacy and procedure level. Products were deemed to have valid quarterly data at the individual patient level if at least one claim was recorded at the pharmacy or procedure level associated with the product. A raw patient uptake curve could then be created for each of the 38 product launches, representing the quarterly patient counts for each product of interest.

Anomaly Detection and Normalized Uptake Curves

A qualitative review was performed on the 38 product launches identifying anomalies in the curves that were unlikely to be directly associated with the launch. The objective of this analysis was to identify the uptake of products in the absence of confounding factors such as supply issues or data gaps, among others. Therefore, using time series statistical forecasting techniques, the raw uptake curves for the 38 product launches were transformed into normalized curves removing the influence of potential confounding factors. To facilitate curve comparison, each update curve was forced to a scale of 0% to 100%, where 100% was considered “peak” share using a qualitative definition of the patient counts increasing by less than 2% per quarter.

Results

As shown in Figure 3, analyses revealed that rare and ultra-rare disease treatments experience significantly faster adoption compared to drugs in non-rare diseases. The vast majority follow an 'R-shaped' uptake curve, characterized by steep early adoption, rapid growth, and a relatively short time to peak. On average, 40%, 60%, and 80% of peak patient volume was achieved at years 1, 2, and 3, respectively, with peak penetration evidenced by year 5. This is in contrast to results from analyses in non-rare diseases conducted by the authors, which show slower rates of adoption and a longer time to peak, with 20% of eventual peak share penetration achieved no earlier than 1 year after launch, a peak penetration attained, on average, after at least 6 years post-launch.

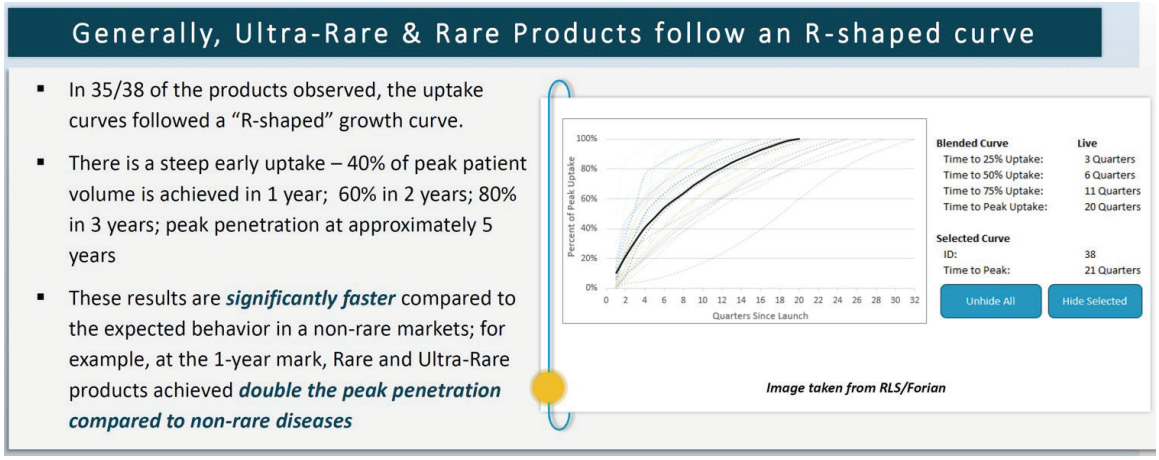


Figure 3: Type of uptake curve for rare disease products

Figure 4 shows that products that took the longest time to peak (≥25 quarters compared to the average of 20 quarters) were launched and marketed by small-medium companies. The one exception in the figure below is currently marketed by a large company but was launched by a small-medium company.

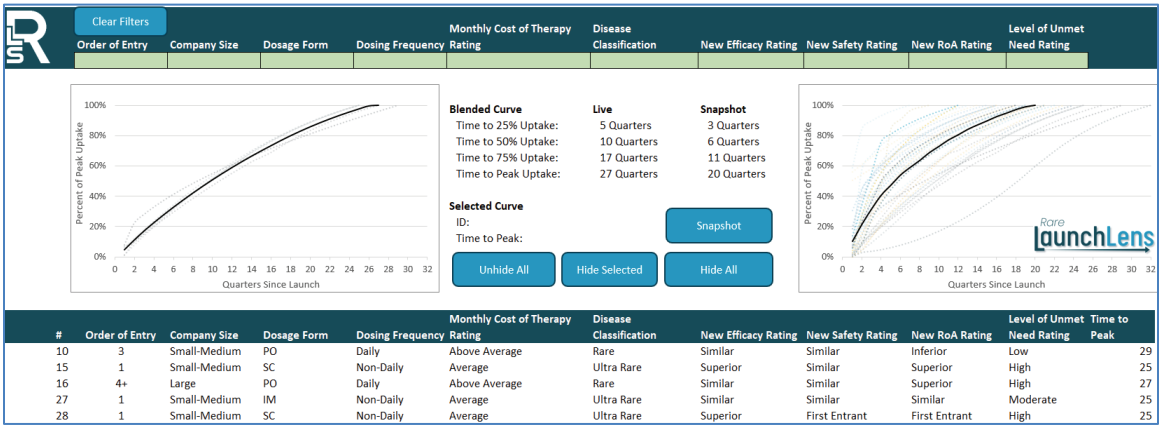


Figure 4: The longest uptake curves were exhibited by small companies

Figure 5 shows that new treatments with superior efficacy compared to drugs already marketed at the time of launch experienced faster uptake and reached peak approximately 3 quarters earlier (19 vs 22). What was observed was that in higher unmet need disease states, efficacy was the main driver of uptake. In contrast, in lower unmet need markets, factors such as route of administration and safety played a more significant role in driving the speed of uptake.

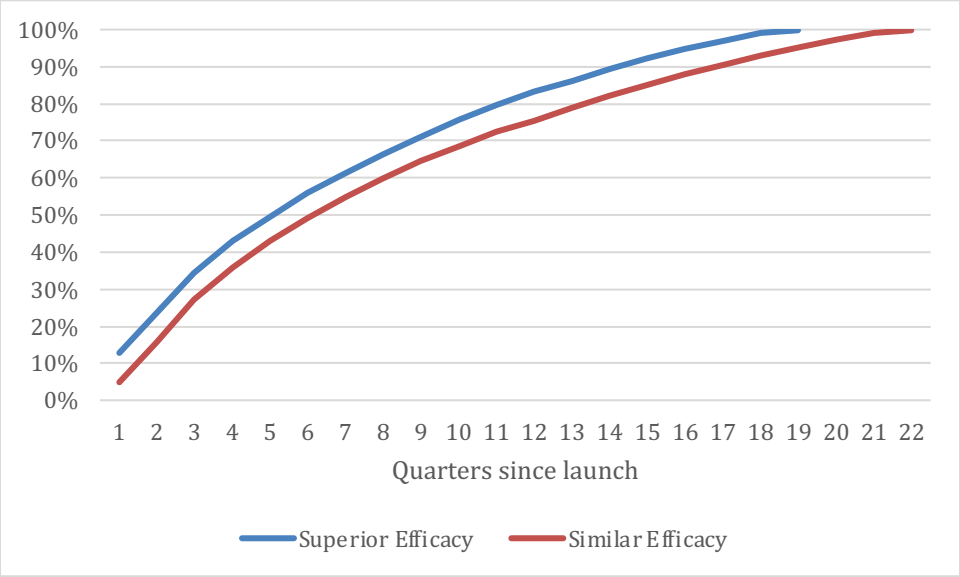


Figure 5: Patient uptake curves by efficacy rating

Discussion

Results from these analyses provided evidence of several factors potentially influencing new products' uptake in rare diseases. Differences by attributes and general direction of findings were in line with expectations. Order of entry, efficacy rating, level of unmet need, and company size (likely a surrogate for launch marketing and prowess) were highly correlated to attained peak patient volume.[i]

Confirming preliminary expectations, there was a very strong association between efficacy and unmet need (i.e. only suboptimal or no effective treatment options were available at the time of launch). Conversely, in indications where treatment options existed at the time of launch, therefore lowering the level of unmet need, more products with similar efficacy relative to prior treatments were available. Unmet need was a significant driver of patient volume uptake; in indications with greater unmet need, improvements in product efficacy, regardless of their magnitude, were associated with a marked increase in speed of uptake compared to similar efficacy products.

Results from this analysis also highlighted the importance of other product attributes as potential drivers of product uptake. As unmet needs become progressively satisfied due to the availability of treatments providing at least minimally acceptable efficacy, other factors

[i] In subsequent analysis not reported in this paper, the authors have identified that these variables, as well as others were statistically significant predictors of attained peak patient volume.

including order of entry, company size, and disease classification (rare vs ultra-rare) were associated with speed of uptake.

As in non-rare disease markets, order of entry (OOE) in rare diseases may influence the rate of product uptake. A small, but detectable OOE effect was found in these analyses. Nonetheless, since this attribute was highly correlated with efficacy, these results may be at least partially confounded, suggesting that later products may be more likely to have a similar or marginally superior efficacy profile, therefore modifying the potential effect of OOE.

Another interesting finding from these analyses was the potential influence of company size on the uptake of rare disease products. Compared to small-to-medium sized companies, products marketed by large companies reached peak patient penetration significantly earlier. Several factors may explain this difference, most likely related to promotional investment levels, increased geographical reach, and prior experience in rare disease product marketing.

Finally, analyses also showed differences in uptake between rare and ultra-rare disease products. In line with expectations, products launched in ultra-rare indications displayed a faster uptake rate than that of rare disease products. As was the case with other attributes, these differences may be due to some confounding with unmet need but may be also reflective of companies' ability to reach more limited patient populations.

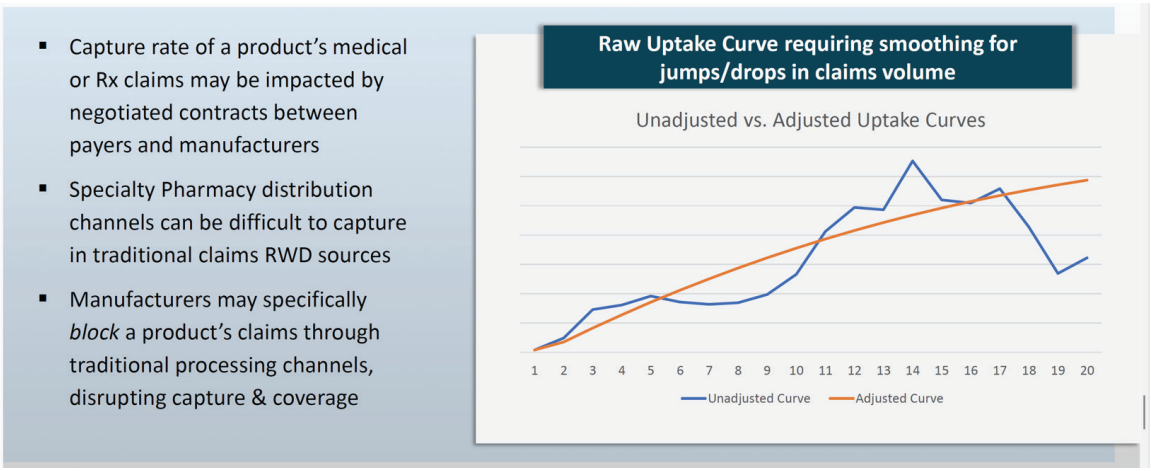
Challenges and Caveats

Several challenges were identified during the analysis. As with all transactional RWD data sources, the prescription claims data included in this analysis were limited to patient and treatment characteristics required to complete a healthcare revenue cycle for prescription drugs. When modeling outcomes using healthcare claims data, there is potential for missing data, misclassification, or unmeasured confounding to threaten the internal and external validity of the analysis. These threats are minimized in CHRONOS due to the linked nature of the data source, combining patient data from multiple sources; CHRONOS combines prescription data captured from healthcare payers, clearinghouses, switches, and other sources; Not only is missingness in patient treatment reduced but this expansive data source represents a large proportion of the US population reducing issues of generalizability in rare disease cohorts.

One limitation was that the data in the analytical sample were restricted to prescription claims captured within the Forian CHRONOS ecosystem. Forian/CHRONOS does contain the full claim adjudication lifecycle and future data analysis is expected to include this additional dimensionality. This may have led to gaps in the data if certain prescriptions were processed outside traditional claims channels. Additionally, specialty pharmacy distribution channels and the possibility of some manufacturers blocking the claims of certain products posed challenges for data capture.

The example (see Figure 6) shows how a product's medical or Rx claim may be impacted by negotiated contracts that non-insiders may not have visibility to. To correct for these types of data anomalies, each product curve was individually reviewed, using "best practice" approaches to understand trend breaks. As shown in this example, a simple examination of the claims versus "reported net sales" suggests that the trend break is artificial, likely due to a data capture issue. The authors applied traditional trending tools to adjust curves and peak patient volume assessment.

Figure 6: Challenges of using real-world data



Despite these limitations, the analysis was strengthened by the expansive nature of the CHRONOS dataset, which includes patient data from a wide range of healthcare providers and payers, helping to reduce issues related to missing data or generalizability.

Conclusion

The study identified several factors that may influence the speed of product uptake in rare disease markets, including efficacy, company size, order of entry and unmet need. The insights gained from this analysis have important implications for commercialization strategies and business development decisions.

Increased accuracy in forecasting rare and ultra-rate uptake curves and time-to-peak adoption can improve the financial valuation of new products and assist in "Go/No-Go" decisions. Using inadequate uptake curves and model assumptions may significantly and negatively impact the NPV/ROI market valuations. Understanding the appropriate rate of adoption (i.e. uptake curve) as opposed to only the final peak share, as significant implications for commercialization, business development and M&A decisions. Furthermore, understanding the drivers of adoption of rare disease products can inform public health policy and patient assistance programs.

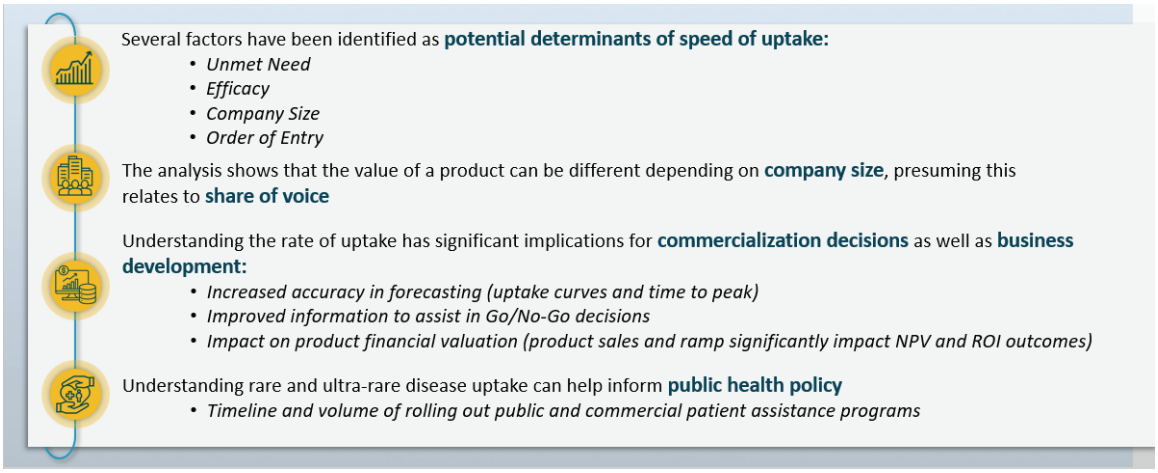


Figure 7: Conclusions of the analysis

Next Steps

Future research should explore how statistical predictors of uptake differ between rare and non-rare diseases. Additional variables, such as route of administration, safety rating, and cost of therapy, may be incorporated into analytical and predictive models. Further analyses should also explore potential differences in adoption dynamics among non-rare therapeutic areas, as well as between acute and chronic conditions.

References

1. Orphan Drug Act, No. 97–414, 96 Stat. 2049 (1983).
2. Jackson C, Telang S. Global Rare Disease Drug Development: Understanding and mitigating its complexities. 2016.
3. Haendel M, Vasilevsky N, Unni D, Bologa C, Harris N, Rehm H, et al. How many rare diseases are there? Nat Rev Drug Discov. 2020;19(2):77-8.
4. The Lancet N. Rare neurological diseases: a united approach is needed. Lancet Neurol. 2011;10(2):109.
5. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet. 2020;28(2):165-73.
6. Ferreira CR. The burden of rare diseases. Am J Med Genet A. 2019;179(6):885-92.
7. Chung CCY, Hong Kong Genome P, Chu ATW, Chung BHY. Rare disease emerging as a global public health priority. Front Public Health. 2022;10:1028545.
8. Faye F, Crocione C, Anido de Pena R, Bellagambi S, Escati Penaloza L, Hunter A, et al. Time to diagnosis and determinants of diagnostic delays of people living with a rare disease: results of a Rare Barometer retrospective patient survey. Eur J Hum Genet. 2024;32(9):1116-26.
9. Groft SC, Posada M, Taruscio D. Progress, challenges and global approaches to rare diseases. Acta Paediatr. 2021;110(10):2711-6.
10. Fermaglich LJ, Miller KL. A comprehensive study of the rare diseases and conditions targeted by orphan drug designations and approvals over the forty years of the Orphan Drug Act. Orphanet J Rare Dis. 2023;18(1):163.
11. Health TLG. The landscape for rare diseases in 2024. Lancet Glob Health. 2024;12(3):e341.
12. Miller KL, Lanthier M. Orphan Drug Label Expansions: Analysis Of Subsequent Rare And Common Indication Approvals. Health Aff (Millwood). 2024;43(1):18-26.
13. Office USGA. GAO-25-106774 Rare Disease Drugs. 2024.
14. Administration USFaD. 9 Things to Know About CDER's Efforts on Rare Diseases [Available from: https://www.fda.gov/drugs/things-know-about/9-things-know-about-cders-efforts-rare-diseases.

About the Authors

Jerry Rosenblatt is the founder and CEO of Rosenblatt Life Science Consultants. His professional career has been devoted to advancing Marketing Science - the application of scientific methods to the understanding of marketing and management challenges. He has been working with pharmaceutical companies for more than 30 years applying his expertise to assist clients with Forecasting & Opportunity Assessment, Business Valuation and the creation of innovative marketing education and learning & development programs. He is the former Head of the Global Forecasting & Opportunity Assessment consulting group of IMS Health (IQVIA). Jerry holds an MBA and PhD (Marketing Science) and is a recipient of the PMSA Lifetime Achievement Award.

Adrian Ghenadenik is an expert in forecasting and opportunity assessment of pharmaceutical products. He has managed assessments across all major therapeutic areas, as well as rare diseases, including in-line brand forecasts, new product assessments, in-licensing opportunity evaluation and strategic life cycle management. Previously, he was a Principal with IMS Consulting, leading a worldwide team of more than 30 consultants in the areas of forecasting methodology, market assessment and forecasting training. Adrian holds a Ph.D. in Public Health and a M. Sc. in Community Health from University of Montreal.

Rhys Rosenberg is an Associate Technology Manager at RLS Consultants, where he leverages cutting-edge technologies to create accurate forecasts and streamline processes. With a strong background in data analytics, process optimization, and technology integration, Rhys specializes in developing innovative strategies to answer complex questions and deepen analytical insights.

Ashwin Anand is an Associate Director in the Commercial Services team of FORIAN Inc, working extensively with robust Real-World Data products towards the end delivery of analytic solutions to our clients. Ashwin has worked in the Real-World Data space for multiple years, combined with his previous experience as an Epidemiologist. Prior to joining Forian, Ashwin worked at Decision Resources Group/Clarivate and a Center for Medicare and Medicaid Services (CMS) federal contractor. He holds an MPH in Epidemiology from Boston University.

Measuring the Impact of Tech-Driven Specialty Pharmacy Collaboration

James Battaglia, Claritas Rx

Introduction

This paper embarks on an exploratory journey into the realm of tech-driven specialty pharmacy collaboration, with a primary focus on patients navigating the complexities of specialty and rare diseases. Currently inundated with manual operations characterized by delays and nebulous accountability, our aim is to shed light on the potential of a more tech-centered approach. Drawing upon CRM system pathways, we delve into the impact and benefits that technology could offer to specialty brands, especially in optimizing fill rates.

Background

The current state of care coordination often is slow, demonstrating a reactive rather than proactive approach, particularly for at-risk patients. This delay can largely be attributed to the overreliance on manual processes such as emails, phone calls, and meetings that require complex interchanges among numerous healthcare partners. Such matrixed interactions contribute to blurred accountability, tardy responses, and hard-to-track interactions.

Nonetheless, an avenue of improvement lies in the advent of a collaboration portal specifically designed for real-time surveillance and execution of the next-best action in patient support. A Customer Relationship Management (CRM) system integrated across various manufacturer patient service teams could allow strengthened management of all partners as extensions of a support center, enabling timely interventions to uphold at-risk patients before they discontinue treatment.

The beneficial impacts of such tech-driven specialty pharmacy (SP) collaboration are quantifiable. Typically, an inverse correlation exists between the time taken to fill a prescription and the actual Fill Rate among specialty brands. The longer the patients wait to fill their prescriptions, the lower the probability of them commencing treatment.

Understanding the interaction dynamics of patient support throughout the treatment journey could amplify the success rate of getting and retaining patients on specialty brands.

Research focusing on interventions within the customer demographic reveals that traditional patient management support is advantageous only to hub patients. However, applying an intelligent case coordination CRM for real-time SP collaboration enhances the management of complicated cases far more proficiently than a standard patient management support system. Such an approach positively influences SP Fill Rates, with improvements ranging from 5-20%.

This paper is intended for those interested in patient management systems, operational effectiveness, and strategic intervention planning. It presents a distinctive opportunity to explore the complexities and nuances of different patient management support mechanisms and their applicability in a real-world context.

Methodology

To examine the efficacy of intervention strategies and their impact on brand performance, we employed a multi-faceted analytical approach focused on two main areas: Time to Fill and Fill Rate.

Initially, we evaluated two brands across two distinct brand groups (Customer A/ Brand A and Customer B/Brand B). The primary performance indicators were the quantitative metrics of Time to Fill and Fill Rate. These metrics were compared for both Hub and SP patients, both in scenarios where interventions were implemented and where they were not. This comparative analysis provided insight into the relative effectiveness of our interventions.

The study utilized a matched cohort design, comparing intervention and non-intervention groups across similar patient demographics and disease states. This approach helped ensure that differences observed in Fill Rate and Time to Fill were more likely attributable to the intervention rather than pre-existing patient characteristics.

Moreover, a control group was utilized to enhance our findings' robustness. By analyzing the Time to Fill Rate in relation to this control group, we could better discern whether the observed changes could be attributed to our interventions or whether they were a result of confounding variables.

Lastly, we sought to assess the CRM utilization among SP partners. This customer relationship management tool was evaluated regarding response rate and the speed at which the interventions were implemented. Considering these metrics, we could better understand SP partners' receptiveness to the interventions and ability to execute them effectively.

Results

The study is divided into two core areas based on two business models, Customer A/Brand A and Customer B/ Brand B. Each model displays a different approach toward patient management, providing an in-depth comparison of typical intervention without smart case coordination and enhanced intervention with smart case coordination.*

For Customer A/Brand A, the paper examines a Hub with Enhanced Intervention and an SP with Typical Intervention. These two operational approaches are compared, providing valuable insights into their effectiveness and efficiency in patient management, as well as their influence on Fill Rates and Time to Fill. (see fig 1)

Moving on to Customer B/Brand B, both the Hub and SP operate under Enhanced Intervention. This unified model is intricately studied to provide a comprehensive understanding of the impact and benefits of Enhanced Intervention across all levels of patient management, as well as their influence on Fill Rates and Time to Fill.

* The smart case coordination portal augments collaborative efforts and expedites the management of cases by enabling all involved parties to access consolidated data. This capability facilitates the prompt resolution of patient treatment challenges, with 85% of users obtaining responses within a 24-hour period, thereby ensuring timely support.

Figure 1: A Comparative Evaluation of Hub-Based Typical Interventions Versus Enhanced Smart Case Coordination Interventions.

Customers	Hub	SP
Customer A (Brand A)	Enhanced Intervention	Typical Intervention
Customer B (Brand B)	Enhanced Intervention	Enhanced Intervention

Evaluated customers with typical patient management support in Hub and varied SP patient management support (typical intervention without smart case coordination to enhanced intervention with smart case coordination)

Customer A made use of Brand A's enhanced intervention support for the Hub and standard intervention for SP.

- Hub = Enhanced Intervention
- SP = Typical Intervention

Impact in this analysis is defined as the percentage change in Fill Rate for patients who received an intervention or not compared to their control cohorts. The study confirms that an enhanced intervention method for hub patients positively impacted prescription Fill Rates, but only for prescriptions with a fill time of 30 days or more. This led to a 13% increase in Fill Rate for hub patients.

However, traditional intervention methods at SPs did not yield significant changes, highlighting the need for technology-driven enhanced collaboration. Traditional SP interventions typically involve more manual outreach efforts such as phone calls to patients to remind them to pick up prescriptions, passive prescription status tracking, and basic adherence counseling. These methods often lack real-time data integration, limiting their ability to proactively address patient barriers to fulfillment. Without automated workflows and dynamic case escalation, many high-risk patients are left without timely intervention, resulting in lower overall Fill Rates.

Figure 2: Assessing the Impact of Hub (Enhanced Intervention) and SP (Typical Intervention) on Customer A's engagement with Brand A

Time to Fill Bin	Fill Rate	Hub Patient Counts	Hub Patients – No Intervention	Hub Patients – Intervention	Impact of Hub Patient Fill Rate	SP Patients – No Interventions	SP Patients – Interventions	Impact of SP Fill Rate
0 - 15 Days	78.2%	1408	15%	85%	0%	82%	18%	0%
15 - 30 Days	73.3%	681			0%			0%
30+ Days	51.2%	737			13%			0%

Takeaway: Brand A exhibits a common pattern where the longer it takes patients to start on a specialty brand, the less likely they are to do so; only hub patients benefit materially from intervention and support. Implementing an "enhanced" intervention method for hub patients impacted prescription Fill Rates, but only for prescriptions with a fill time of 30 days or more. This led to a 13% increase in the Fill Rate for such cases

Enhanced SP Intervention Improves Fill Rates

For Customer B/Brand B, both the Hub and SP operate under Enhanced Intervention.

- Hub = Enhanced Intervention
- SP = Enhanced Intervention

On the SP side, enhanced interventions refer to a more proactive approach leveraging CRM-driven workflows to identify patients at risk of discontinuation. This facilitates more frequent and timely patient services interventions across care teams leading to improved efficiency, as well as increased fill rates and adherence, compared to what is typically done within the patient services center. Examples of such interventions include providing appeals support for prior authorization denials, assisting patients with financial hardship through copay card acquisition, engaging with SPs to address scheduling delays and operational inefficiencies, and interacting with HCPs to mitigate missed fills and drive adherence.

The Customer B/Brand B Fill Rate by Time to Fill Cohort showed a 20% increase in Fill Rate for prescriptions taking more than 30 days to fill and a 5% increase in Fill Rate for prescriptions filled between 15-30 days. The study demonstrates that CRM-driven smart case coordination enhances specialty pharmacy engagement, reduces fulfillment delays, and improves overall adherence rates especially for more high-risk patients. This underscores the role of technology as a critical enabler of specialty patient management.

Figure 3: For Customer B/Brand B, both the Hub and SP operate under Enhanced Intervention.

Time to Fill Bin	SP Patients – No Interventions	SP Patients – Interventions	% With Intervention	Control Cohort Fill Rate	Intervention Cohort Fill Rate	Impact of SP Fill Rate
0 - 15 Days	324	101	23%	69%	67%	0%
15 - 30 Days	92	105	53%	56%	61%	5%
30+ Days	18	84	82%	28%	48%	20%

Intervention in more high-risk patients than is typical

Takeaway: Our research presents an in-depth evaluation of the impact of tech-driven specialty pharmacy collaboration, focusing on utilizing the Platform. This research investigates the frequency of interaction between PBM-owned and independent specialty pharmacies (SPs) and the platform across various programs. It is based on our findings from five major SP partners - Biologics, Accredo, Kroger, CVS, Caremark, and Optum.

Figure 4: PBM-owned and independent SPs engage with the platform across programs.

SP Partner	Total Number of Inquiries	Total Number of Responses	Response Rate	Median Response Time (Days)
Partner 1	523	464	89%	< 1 Day
Partner 2	2836	2370	84%	< 1 Day
Partner 3	553	441	79%	1 Day
Partner 4	4461	3426	77%	< 1 Day
Partner 5	219	168	77%	2 Days

Takeaway: An exploratory study was performed on the utilization and interplay of the CRM with SPs. This research furthermore probed potential correlations between the timely responses from client-associated SP partners and the boosted engagement of Field teams towards augmenting patient interventions, along with their consequent outcomes. There was improvement in Response Rate and median response time. Results indicate that over 81% of responses occur within a day, demonstrating the effectiveness of this tool in driving FRM engagement and enhancing patient intervention.

Discussion

The findings underscore the critical impact of technology-enabled real-time case coordination in improving Fill Rates among SP patients. The study reinforces that traditional patient management frameworks often fall short in addressing the complexities of specialty pharmacy dynamics. However, integrating CRM-driven intervention tools with real-time SP collaboration has demonstrated measurable efficiency gains. Customer B/Brand B's adoption of this approach significantly improved the resolution of challenging cases, outperforming the intervention strategies employed by Brand A.

A key observation from this study is that while hubs have traditionally been the primary focus for patient interventions, specialty pharmacy engagement has remained largely reactive and inconsistent. Traditional SP interventions, which rely on manual outreach efforts such as reminder calls, status tracking, and general adherence counseling, are often insufficient in addressing the nuanced barriers that prevent patients from filling their prescriptions. Without a coordinated, technology-driven approach, these interventions lack real-time visibility into patient behaviors and fail to proactively mitigate prescription delays.

Enhanced interventions facilitated through CRM-driven smart case coordination offer a fundamental shift in how specialty pharmacy collaborations function. The ability to automate workflows, escalate high-risk cases, and provide SPs with structured engagement strategies significantly enhances patient support outcomes. Furthermore, the ability to track and analyze

intervention effectiveness at both the individual patient level and across broader populations ensures that intervention strategies can be continuously refined to maximize impact.

Ultimately, these findings reinforce that technology-driven solutions not only improve Fill Rates but also optimize operational efficiency for all stakeholders involved in specialty medication fulfillment. For example, a swift response (within 1-2 days) from client SP partners was found to foster the FRM team's functionality, facilitate improved engagement with HCPs, and provide quicker support for the patients. These findings demonstrate the potential benefits of adopting innovative, collaborative solutions in improving patient outcomes.

The data demonstrate that enhanced interventions are most effective when they are proactive, targeted, and supported by digital infrastructure that facilitates seamless communication across all touchpoints of the patient journey.

Conclusion

This study demonstrates that the patient care strategies employed by Brand B, particularly through CRM-driven Smart Case Coordination, significantly enhance real-time collaboration and improve patient outcomes. By enabling faster and more effective engagement between specialty pharmacies, hubs, and healthcare providers, this technology ensures that patients receive timely support, reducing delays in therapy initiation and improving adherence.

The ability of CRM-enabled solutions to drive rapid responses from SP partners highlights their potential to strengthen HCP engagement and streamline patient support. As healthcare systems continue to evolve, the integration of advanced digital tools that facilitate proactive intervention will be critical in optimizing specialty pharmacy operations and ensuring seamless patient care.

More broadly, this study reinforces the necessity of incorporating cutting-edge technologies into healthcare ecosystems. The rapid response mechanisms and structured workflows leveraged by Brand B demonstrate how technology can bridge gaps in patient care coordination and drive measurable improvements in therapy access. By fostering collaboration between manufacturers, SPs, and healthcare providers, organizations can create a more efficient, responsive, and patient-centered specialty pharmacy landscape. Looking ahead, continued investment in digital infrastructure and strategic partnerships will be essential for shaping the future of specialty care and improving patient outcomes.

About the Author

James Battaglia is the Director of Professional Services at Claritas Rx, where he leverages over ten years of experience in data science and analytics. His expertise lies in software subscription services, program management, and formulating strategies for revenue growth. James is particularly passionate about optimizing processes, developing key performance indicators (KPIs), and measuring performance to drive scalable and sustainable business outcomes.

Iterative Causal Segmentation*

Filling the Gap between Market Segmentation and Marketing Strategy

Kaihua Ding,[1] Jingsong Cui,[1] Mohammad Soltani,[1] and Jing Jin[1]
[1]AZ Brain, AstraZeneca PLC

Abstract

The field of causal Machine Learning (ML) has made significant strides in recent years. Notable breakthroughs include methods such as meta learners [8] and heterogeneous doubly robust estimators [3] introduced in the last five years. Despite these advancements, the field still faces challenges, particularly in managing tightly coupled systems where both the causal treatment variable and a confounding covariate must serve as key decision-making indicators. This scenario is common in applications of causal ML for marketing, such as marketing segmentation and incremental marketing uplift. In this work, we present our formally proven algorithm, iterative causal segmentation, to address this issue.

1 Motivation

The integration of machine learning into market segmentation has significantly transformed the development of marketing messages and strategies. However, categorizing individuals into rigid market segments such as 'loyalists' or 'dabblers' fails to account for the dynamic nature of consumer circumstances, potentially leading to ineffective marketing and wasted resources. This underscores the limitations of relying solely on traditional market segmentation for marketing actions, a challenge initially highlighted by Wendell R. Smith in the 1950s [7]. Despite its innovative approach, market segmentation's static methodology struggles to capture evolving consumer behaviors, risking oversimplification and ineffective strategy development.

The emergence of causal inference in marketing provides a solution by identifying the causative factors behind consumer behaviors, enabling the development of predictive and effective marketing strategies with methods like Uplift Trees and Meta Learners [6, 8, 11]. However, these approaches often overlook the complexity that arises when market segmentation becomes an intertwined part of the promotional market, acting both as a significant confounder and a desired output for marketing purposes, especially in contexts like pharmaceutical call planning and broadcasting media. Causality analysis alone cannot address it as both a confounder and an output.

To tackle these challenges, we propose the iterative causal segmentation algorithm, which merges causal inference with market segmentation to surmount their individual limitations. This method not only provides a nuanced understanding of consumer behaviors but also amplifies the effectiveness of marketing efforts across various segments. Nonetheless, in marketing scenarios where the interdependency between segmentation and causality analysis poses a distinct challenge, leading to a cyclical problem where each influences the outcome of the other, our proposed solution aims to harmonize these methodologies. By leveraging their strengths, we seek to refine marketing strategies effectively.

*Disclaimer: The opinions expressed in this presentation are those of the presenters and do not necessarily reflect the views of AstraZeneca. The analyses presented in this presentation are based on Uber's open-source CausalML GitHub repository data and do not represent AstraZeneca data.

2 Iterative Causal Segmentation

2.1 Average Treatment Effect (ATE)

A key challenge for marketing professionals concerns generating a sales uplift from promotional campaigns, quantified by:

$$\text{Expected uplift gain from applying promotion} = E(Y^{A=1}) - E(Y^{A=0}), \tag{1}$$

where:

- A represents the treatment, i.e., the promotional campaign.
- Y is the outcome, indicating the effect of the campaign.

The difference between these two expected values represents the expected uplift per person due to the promotion. Equation 1 represents the expected average purchase uplift if the promotion is applied. This expected uplift is also known as the Average Treatment Effect (ATE) in causality analysis. The Average Treatment Effect is defined as:

$$\text{ATE} = E(Y^{A=1}) - E(Y^{A=0}) \tag{2}$$

This formulation allows us to quantitatively assess the overall impact of the treatment across the entire population. Additionally, recently developed causal machine learning techniques are capable of attributing uplift gain to individual samples [2, 8] by estimating the conditional average treatment effect.

2.2 Conditional Average Treatment Effect (CATE)

The Conditional Average Treatment Effect, which can also be understood as the expected individual treatment effect (ITE) conditional on covariates X, is defined as:

$$\text{CATE} = E(Y^{A=1}|X) - E(Y^{A=0}|X) \tag{3}$$

where:

- A represents the treatment, i.e., the promotional campaign.
- Y is the outcome, indicating the effect of the campaign.
- X represents the covariates or features of the individual units that might affect how the treatment impacts the outcome.

CATE provides a more nuanced understanding of how the effect of the treatment varies across different segments of the population based on the covariates X. It allows for the estimation of how the treatment effect differs among individuals or specific groups within the population, enabling targeted interventions.

2.3 Causal Graph

A causal graph, often used in the context of causal inference and statistics, is a graphical representation that models the causal relationships between variables. It is a type of directed graph where nodes represent variables (e.g., events, conditions, or quantities), and edges represent causal effects from one variable to another. This concept is foundational in understanding causal inference, allowing researchers to visually represent and analyze the cause-and-effect relationships within a system.

We can express the relationship among promotion application A, marketing segmentation S, covariates X_i, and promotion outcome Y as the following causal graph.

The causal graph provides an illustrative relationship among treatment A, covariates X, and outcome variables Y [4]. A causal graph is a product of a hypothesis and rationalization, rather than ground truth. Regardless, the causal graph drawn according to empirical knowledge needs to be verified through sensitivity

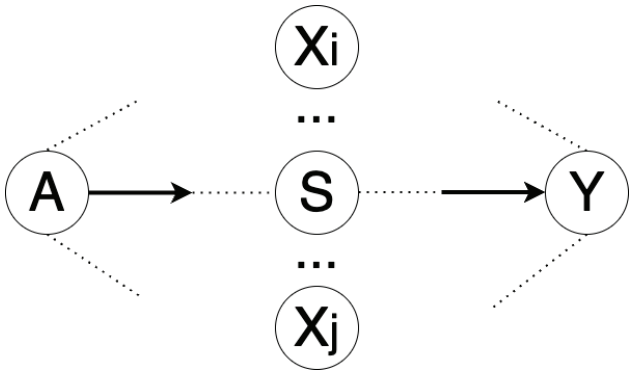


Figure 1: Potential causal graph, where promotion is A, covariates are X_i, promotion outcome is Y, and marketing segments S is a confounder.

analysis. Rather than focusing on how to come up with the best causal graph, the goal of drafting a causal graph should be to guide the collection of covariates and the necessary control variables for covariate matching. In the authors' opinion, establishing a causal graph is helpful but not the most critical aspect of causality analysis. The causal graph should change and needs to change, especially if the causal relationship contains human perspective or ambiguities. It's the overall reliability in the sensitivity analysis that truly matters as the end result.

In addition, the main usage of a causal graph is to assist users in how to control for confounders or covariates to ensure that the causality analysis is valid, e.g., backdoor and frontdoor path criteria [5]. For simplicity, we use the disjunctive criterion [9] in this paper and we perform sensitivity analysis to check the reliability of our causal analysis, which also serves as an assessment for the impact of unobserved confounders and uncertainty quantification.

2.4 The Iterative Causal Segmentation Algorithm

In this section, we illustrate our proposed algorithm, Iterative Causal Segmentation. The key challenge to address is the tightly coupled nature between segmentation and causality analysis. Instead of shying away from the tightly coupled nature of these two computational modules, we propose a joint convergence method that solves for segmentation and causality analysis simultaneously, as shown in Figure 2.

Figure 2 figuratively describes the joint convergence algorithm's workflow, explicitly considering the tightly coupled nature and mutual influence of the two modules. For the purpose of effective uplift behavior segmentation, the causal machine learning module will generate useful uplift incremental estimation, which we will feed into the segmentation module to produce segmentation that closely reflects the promotion uplift effect. Segmentation results will then serve as input in the causal machine learning, and the overall system is considered converged if the amount of segment movement becomes less than the population size variance determined by ATE variance estimates. This workflow can be more formally defined as follows in Algorithm 1.

Algorithm 1 formally describes Figure 2 in a more concise and concrete pseudocode fashion. Observing Algorithm 1, one might question whether there is any need to go to such great lengths to formulate a new algorithm to solve the causality behavior segmentation problem in marketing. This is also the question that the authors are interested in solving first. Are there alternatives to the simpler segmentation algorithm that serve the same purpose without the need to solve the coupled segmentation and causality system? We can formally prove that such an alternative segmentation method does not exist. The causality uplift segmentation analysis is exclusively determined by the iterative causal segmentation algorithms.

We formalize the proof as the following statement of Causal Segmentation Exclusivity. If this theory holds true, it means our iterative causal segmentation is necessary since segmentation obtained this way is exclusive to the causality.

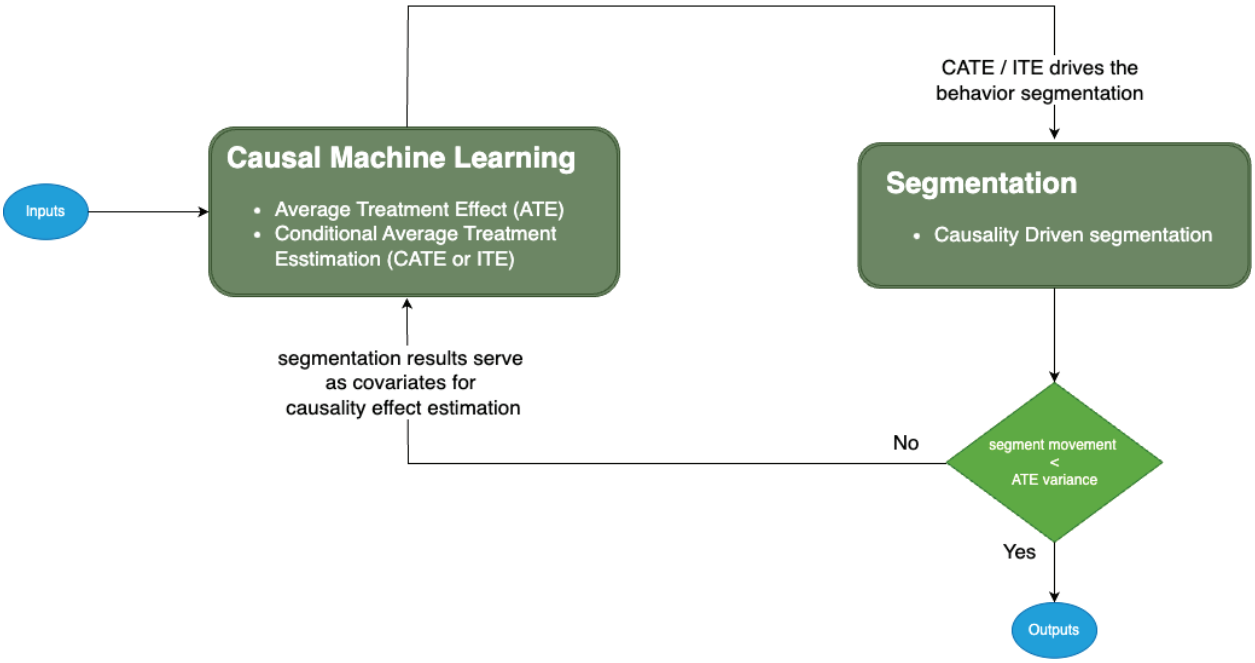


Figure 2: The diagram description on how to use causal machine learning to guide segmentation efforts. The goal is to achieve the joint convergence of the causal machine learning module as well as the convergence of the segmentation module.

Algorithm 1 Causal Machine Learning Driven Segmentation

Require: Treatment variables A, outcome Y, Segmentation S, and covariates X
Ensure: Causal Data Assumptions

Segment Movement ← initial value
ATE Variance ← initial value
while Segment Movement > (ATE Variance × Population Size) **do**
    ATE ← ComputeATE(A, S, Y, X)
    ATE Variance ← EvaluateATEVariance(ATE)
    CATE ← ComputeCATE(A, S, Y, X)
    S ← Segmentation(CATE)
    Segment Movement ← EvaluateSegmentMovement(S)
end while
return S

Theorem 1 (Causal Segmentation Exclusivity)**.** *Let us define the Conditional Average Treatment Effect (CATE), also known as the Individual Treatment Effect (ITE), as follows:*

$$CATE(X) = \mathbb{E}[Y^{A=1} - Y^{A=0} \mid X] \tag{4}$$

where $Y^{A=1}$ represents the potential outcome if treated or received campaign promotion, $Y^{A=0}$ represents the potential outcome if not treated or not receiving campaign promotion, and X is a vector of individual characteristics.

Segmentation, in the context of causal machine learning, is the process of partitioning a population into distinct groups based on their respective CATE estimates. Here, the segmentation function S is

exclusively dependent on $CATE(X)$:

$$S = f(CATE(X)) \tag{5}$$

We hypothesize that no other factors other than CATE influence the segmentation, that is, segmentation is a function of CATE alone.

A formal proof through contradiction is detailed below,

Proof of Theorem 1. Assume for contradiction that there exists another driver D which influences the segmentation such that:

$$S' = f(\text{CATE}(X), D) \tag{6}$$

where S' represents a new segmentation outcome due to the presence of driver D.

If driver D were to affect the segmentation, we would observe a change in the segmentation outcome S without a corresponding change in the CATE estimates. This would contravene the initial assumption that segmentation is exclusively driven by CATE.

Since our operational framework stipulates CATE as the sole driver for segmentation, the supposition of another influencing driver D is invalid. Therefore, changes in segmentation are directly correlated with changes in the CATE estimates, thus affirming that the segmentation is indeed a causal behavior segmentation when CATE is the exclusive driver. ■

The Proof 2.4 formally proves that the iterative causal segmentation is a necessary system to address when the tightly coupled nature between causality and segmentation exists. The significance of this proof lies in its affirmation for the development of such algorithms. Now that we have justified the legitimacy of developing this new algorithm, Algorithm 1, we will evaluate its performance in the next section.

3 Results and Discussion

3.1 Data Sources Disclaimer and Discussion

We examine the performance of Algorithm 1 by applying it to open-source data from the Uber CausalML package [1]. We want to emphasize that all causal machine learning algorithms derive their origin from causality analysis. As a result, all the data assumptions required for performing causality analysis need to be true to ensure the comprehensiveness and validity of the analysis, as outlined in Table 1.

Table 1: Causal Data Assumptions. In the table below, subscript indices denote sample indices, and superscript indicates the treatment/control variable.

Assumption	**Expression**
SUTVA	Each unit sample i's potential outcomes Y_i^A are unaffected by the treatment assignment A_j of any other unit j, where $A = 1$ indicates treatment is assigned and $A = 0$ indicates control.
Ignorability	$Y_i^{A=0}, Y_i^{A=1} \perp A_i \vert X_i$ for all units i, where A_i is the treatment assignment and X_i are covariates.
Positivity	$P(A_i = a \vert X_i = x) > 0$ for all levels of treatment a and covariates x.
Consistency	If $A_i = a$, then the observed outcome Y_i is equal to the potential outcome $Y_i^{A=a}$.

The SUTVA assumption concerns the principle that the treatment received by one unit does not affect the outcomes of any other unit. In other words, the potential outcome for any individual is assumed to be independent of the treatment assignments of all other individuals. The ignorability assumption, also known as the conditional independence assumption or no unmeasured confounders assumption, plays a pivotal role in

the field of causal inference, particularly in observational studies where random assignment to treatment and control groups is not feasible. This assumption is crucial for estimating causal effects from observational data, where the potential for confounding variables is a significant concern. The ignorability assumption allows researchers to control for confounding variables through statistical methods such as regression, matching, stratification, or weighting. By adjusting for a comprehensive set of observed covariates X, one can estimate the average treatment effect (ATE) or the average treatment effect on the treated (ATT) as if the treatment assignment were random, mimicking a randomized controlled trial. The positivity assumption, also known as the overlap or support condition, states that every unit (e.g., individual in a study) has a non-zero probability of receiving each treatment level, given the covariates. The consistency assumption states that the potential outcome of an individual under a specific treatment is equal to the observed outcome if the individual receives that treatment. For marketing campaign application, we provide data processing guidelines on how to achieve these assumptions to ensure the validity of causality analysis.

Table 2: Causal Data Assumptions: How to satisfy these requirements through data processing.

Assumption	**How to achieve through data processing**
SUTVA	Ensure the campaign measurement timeline is short enough before noticeable interferences among units are realized.
Ignorability	Control for covariates to achieve independent treatment assignment.
Positivity	Trimming, weighting, or synthetic control group generation to ensure the population data restricts the causality analysis to the overlap region where control and treatment can both be observed.
Consistency	Ensure the consistency of treatment or note if there are different versions of treatment. For new campaigns, a new causality analysis needs to be performed. This is important for assessing multiple treatment effects.

Once data is pre-processed to avoid yielding biased estimates, we can then apply Algorithm 1. The sample result is provided in Section 3.2.

3.2 Numerical Results

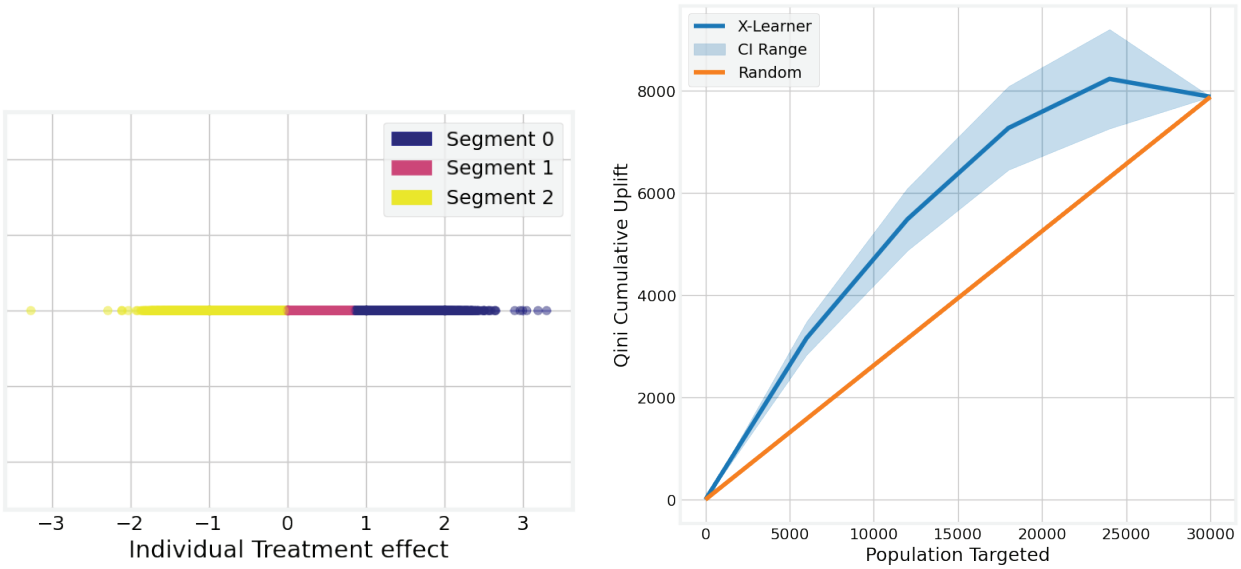
Before we study the numerical results produced by Algorithm 1, we need to ensure the convergence of the algorithm. Below is a sample of the convergence result in Table 3.

Table 3: Converged Result

ATE	SE	SE-ATE Ratio (%)	Movement Precision	Segment Movement
0.507	0.005	1.053	1264.063	1235

In Table 3, the convergence of the causality module produces metrics on ATE, SE, P-Value, SE-ATE Ratio (%), and Movement Precision. The convergence of the overall movement produces the "Segment Movement" that is lower than "Movement Precision" for the overall system to be considered converged. Additionally, we can visualize the converged segmentation results and sensitivity study of the converged results.

Figure 3a shows that the causality is segmented into three segments. For the sensitivity study, we apply the Qini curve measurement [6]. The Qini curve is a performance measurement for uplift modeling, which evaluates the effectiveness of a treatment in a causal inference context. The concept of the Qini curve is



(a) Segmentation Result

(b) Sensitivity Analysis

Figure 3: Segmentation results and sensitivity study result for the converged system described in Table 3

analogous to the Gini coefficient used in economics and the ROC curve used in binary classification models, but it is specifically designed for quantifying the incremental impact of a treatment or intervention. Figure 3b, measured with a Qini curve, shows that the 90% confidence range (CI) is shaded. Overall, it appears that even the bottom envelope of our sensitivity study still shows positive improvement over a random assignment Qini curve.

3.2.1 Simulation Studies and Discussions on KMeans, Propensity Score, and Causal Effect Based Promotion Selection

After causal segmentation converges, we can perform a simulation study following the convergence of Algorithm 1. This simulation study compares causality-based population selection, propensity score-based selection, KMeans-based selection, and a random selection strategy. The relative performance of these four selection strategies is graphed in Figure 4.

Figure 4 is plotted with the population selection percentage on the horizontal axis and the respective cumulative uplift gain on the y-axis. Thus, when 0% of the population is selected, the overall expected promotion gain is 0; when 100%, the selection strategy of any kind no longer matters. Interestingly, the figure shows that the population selected with the causal effect criterion demonstrates the highest overall gain regardless of what percentage of the population is selected for promotions. The KMeans and Random selection curves nearly coincide with each other. Most notably, the promotion target population selected using the propensity score performs even worse than random selection.

Causal Effect For more detail on the causal effect selection strategy, this strategy prioritizes individuals based on their ranked uplift effects. We can intuitively understand the convex curve of the causal effect through Figure 5.

Intuitively, selecting a promotion population based on "Causal Effect" (CATE) in descending order is akin to choosing users with the highest uplift first, represented in the lower right corner of Figure 5. The initial slope of the "Causal Effect" is the steepest because it involves selecting users from a small portion of the population. The "Causal Effect" curve will gradually level to match the slope of the "Random" curve, indicating a transition towards random population selection. Eventually, the "Causal Effect" curve may adopt a negative slope, signaling that users with negative uplift are being selected.

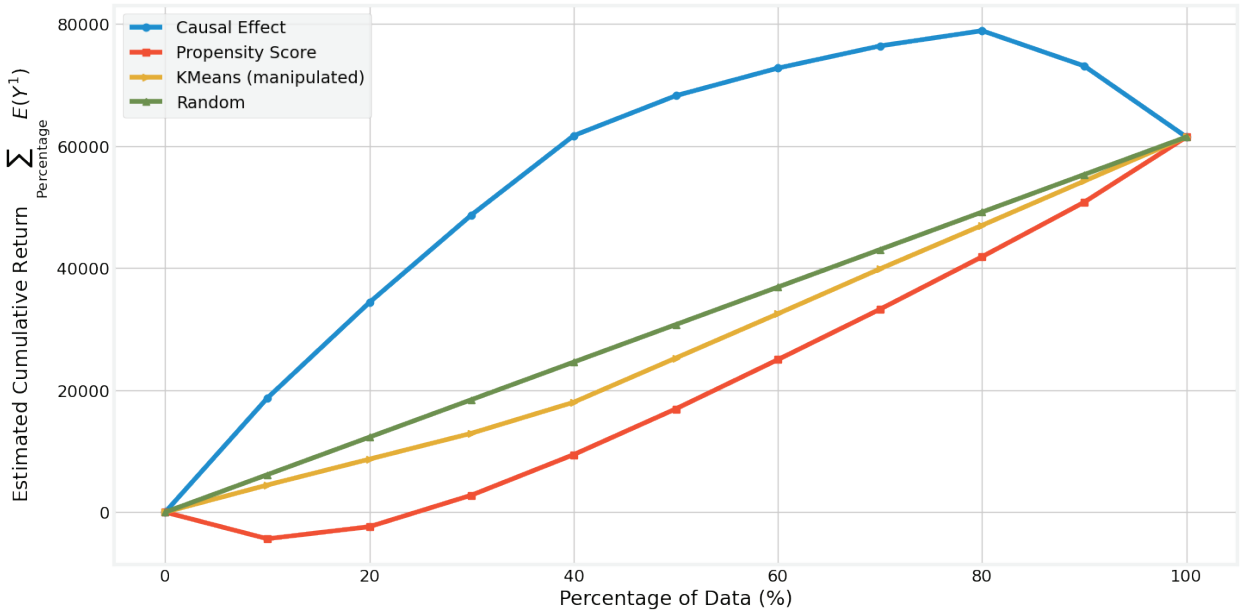


Figure 4: Simulation study comparing the performance among the cumulative return gain of four different promotion population selection strategies. Causal Effect selection is based on the treatment effect. Propensity score selection is based on the propensity, which is regressed from the given X to the propensity of obtaining outcome Y. KMeans is based on clustering results using X. Random selection is a random promotion assignment, whose expected slope is the same as the expected gain, i.e., ATE.

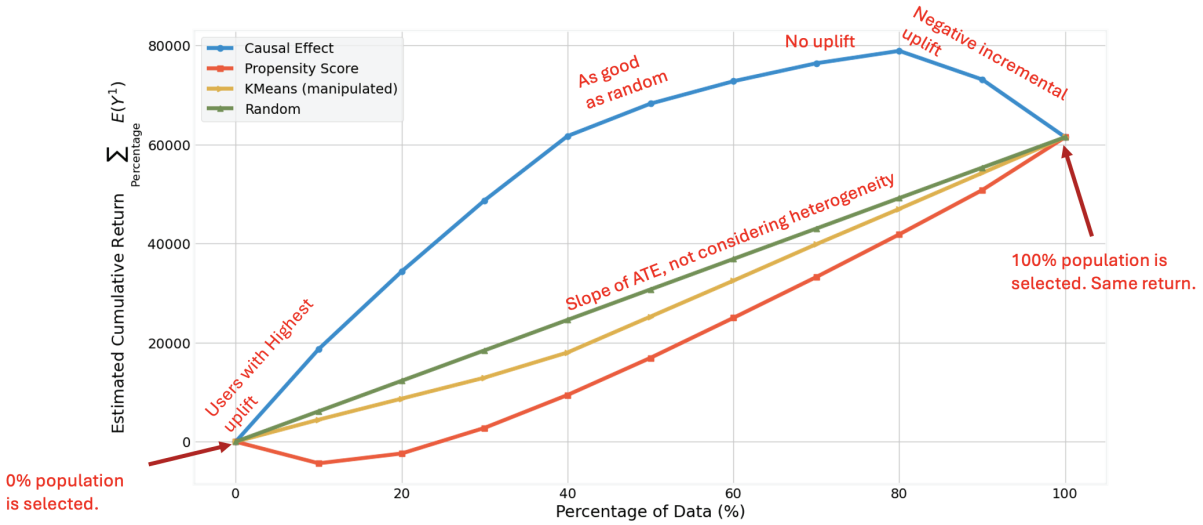


Figure 5: Simulation study comparing the performance in terms of cumulative return gain among four different promotion population selection strategies. Specifically, the causal effect curve is labeled.

K-Means Why Does Pure K-Means Without Causality Segmentation Perform Poorly? As depicted in Figures 4 and 5, the performance of K-Means aligns closely with that of random assignment strategies for promotion allocation. K-Means is an unsupervised learning algorithm that focuses on partitioning datasets into k groups based on feature similarity. It aims to divide the n observations into k clusters, where each observation is assigned to the cluster with the nearest mean, serving as the prototype for that cluster.

The clusters formed by K-Means are based on the mathematical criterion of minimizing the variance within clusters, as measured by the Euclidean distance. This objective does not necessarily align with human intuition or domain-specific, meaningful groupings. Therefore, while not "artificial" in the sense of being random or arbitrary, the clusters may not always correspond to explainable or expected patterns in the data. K-Means does not ensure that the results of clustering will be inherently understandable or match known categories within the data. The algorithm identifies structures based on its mathematical objective, which may or may not coincide with meaningful or recognizable categories to humans.

K-Means is a powerful tool for exploratory data analysis and can uncover intriguing patterns within the data. However, its simplicity and the nature of its objective function mean that it is most effective under specific conditions—namely, when supplemented by domain knowledge, additional context, or other clustering methods for interpreting the results.

Figure 6 showcases a sample of behavior segmentation results commonly utilized by marketing strategists. The characterization of each K-Means segmented segment is based on a posteriori interpretation rather than predictive outcomes. For instance, Segment A in Figure 6 is identified as representing loyal, highly interested customers, while Segment E is categorized as comprising highly critical and cautious customers. This segmentation information is relevant for businesses. However, these interpretations are not directly derived from K-Means; the computation steps of K-Means were not informed by these specific objectives or personas. Thus, the explanation of the segments is somewhat artificial, as illustrated in Figure 6.

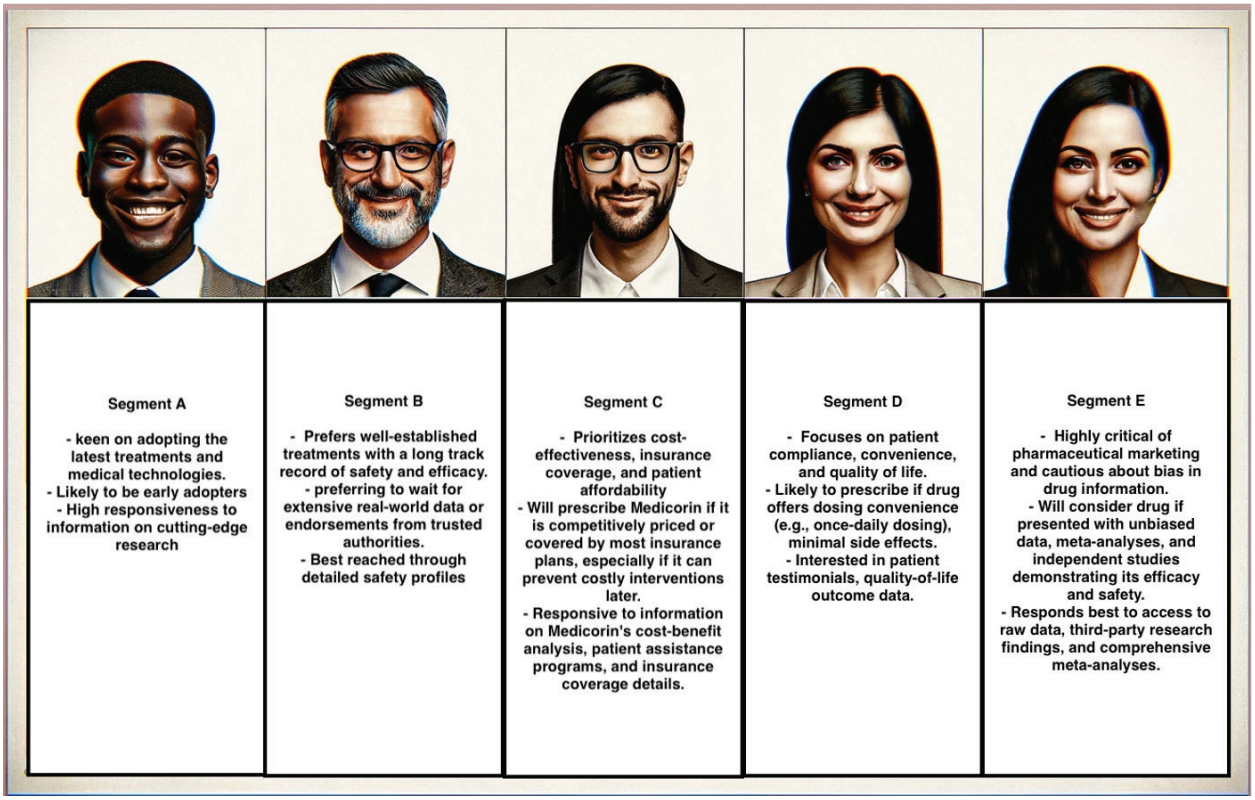


Figure 6: When K-Means segmentation is performed for marketing purposes, the segmentation initially relies on input covariates that represent behavior traits. After segmentation is completed, each segment is then characterized by distinct behaviors.[+]

Propensity Score The most interesting curve in the simulation studies depicted in Figures 4 and 5 relates to the propensity score, which actually performed worse than both the causal effect-based and K-Means-based promotion selection strategies. Propensity models, which assess the relationship between covariates

[+]Disclaimer: Medicorin is a fictional drug name used for illustrative purpose only.

X and the purchase outcome Y, do not inherently aim to answer how to select a population for promotion activities to maximize uplift gain. Essentially, a propensity model only addresses the likelihood of a purchase occurring given X, not how to select individuals for promotional activities to achieve the greatest uplift.

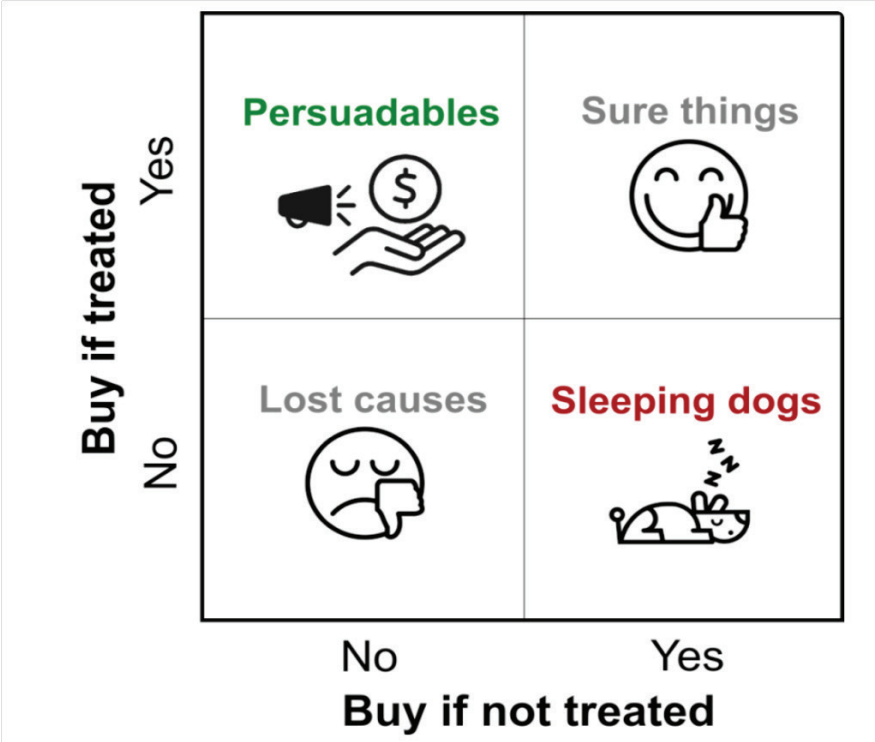


Figure 7: An illustration of quadrants [10]

Figure 7 illustrates that even if someone has a high propensity to purchase, this should not serve as the sole criterion for deciding whether to allocate marketing resources to specific customers. For example, customers identified as 'Sure things', despite their high propensity, are not ideal candidates for spending marketing dollars on. Conversely, just because customers have a low propensity to purchase does not mean they should automatically be excluded from marketing efforts, especially if they fall into the "Lost Causes" quadrant.

Table 4 clearly defines the propensity score as $P(\text{Purchase} \mid \text{No Intervention})$, which is distinct from the objectives addressed by causal effect analysis in Equations 2 and 3.

Table 4: Machine Learning Model Comparison

ML model	**Model tries to answer**	**Business problem we hope to solve**
Propensity model	$P(\text{Purchase} \mid \text{No Intervention})$	Find audiences with: High $P(\text{Purchase} \mid \text{Intervention})$ Low $P(\text{Purchase} \mid \text{Intervention})$
Churn model	$P(\text{Churn} \mid \text{No Intervention})$	Find audiences with: High $P(\text{Churn} \mid \text{Intervention})$ Low $P(\text{Churn} \mid \text{Intervention})$
Response model	$P(\text{Purchase} \mid \text{Intervention})$	Find audiences with: High $P(\text{Purchase} \mid \text{Intervention})$ Low $P(\text{Purchase} \mid \text{Intervention})$

Therefore, neither the churn model $P(\text{Churn} \mid \text{No Intervention})$ nor the response model $P(\text{Purchase} \mid \text{Intervention})$

should be used as strategies for selecting customers for promotions.

However, this does not fully explain why the propensity score selection strategy is the least effective among all promotional target selection strategies depicted in Figures 4 and 5. To clarify, we examine the relationship between propensity scores and CATE, as illustrated in Figure 8.

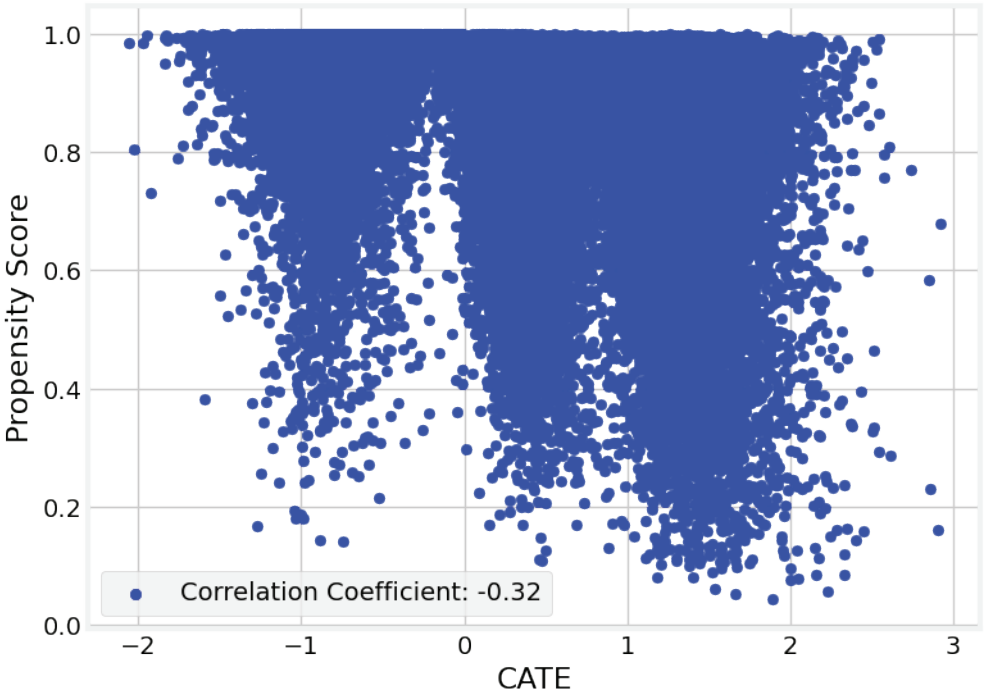


Figure 8: Not only are they not correlated, they are mildly negatively correlated.

Figure 8 shows that propensity scores and CATE are not strongly correlated; in fact, they are weakly negatively correlated. Therefore, propensity scores cannot serve as a substitute for causality analysis in uplift modeling.

3.3 Explainability

Given the exclusivity between causality analysis and segmentation results as outlined in Theorem 1, the explainability of the overall iterative causal segmentation algorithm merits discussion. Although the segmentation algorithm is unsupervised and ordered by thresholding, Algorithm 1 can be elucidated through its causality module. This module records the converged state when all three modules—causality, segmentation, and segment movement—have converged. Techniques like SHAP values can be adopted to provide granular explanations.

Since SHAP values can offer individualized explanations regarding features, as demonstrated in Figure 9, they can be utilized for the causality module of the iterative causal segmentation algorithm (Algorithm 1). Moreover, due to the exclusivity detailed in Theorem 1, the SHAP value explainability for the causality module also extends to the overall explainability of the iterative causal segmentation algorithm.

4 Conclusion

In this paper, we have introduced the iterative causal segmentation algorithm, Algorithm 1, designed specifically for marketing contexts where segmentation strategies play a crucial role in influencing purchase outcomes. This necessitates addressing the tightly coupled system between promotion and segmentation.

We demonstrated the value of this segmentation algorithm by comparing it with other common machine learning models used in marketing settings in Section 3.2.1. Furthermore, we established the exclusivity of

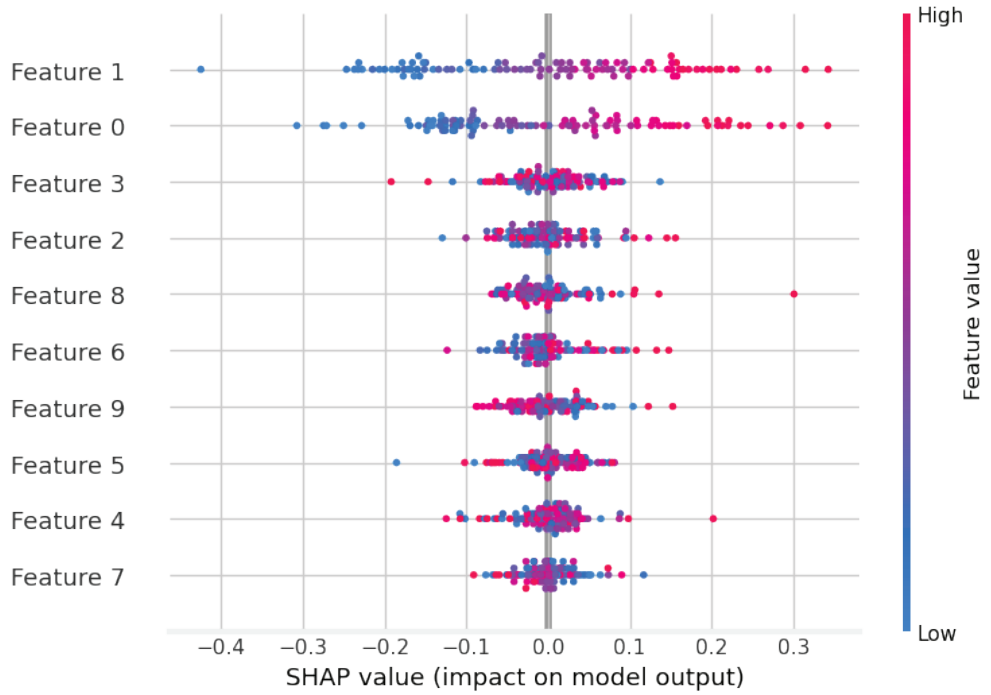


Figure 9: Example of SHAP value explainability for the converged results of Table 3 with a subpopulation of 100 data points. Note: SHAP value computation can be computationally expensive.

the proposed segmentation method, showing that it cannot be easily replaced by other methods, as evidenced in Section 2.4.

References

[1] Huigang Chen, Totte Harinen, Jeong-Yoon Lee, Mike Yung, and Zhenyu Zhao. Causalml: Python package for causal machine learning, 2020.

[2] Behram Hansotia and Brad Rukstales. Incremental value modeling. *Journal of Interactive Marketing*, 16.3, 2002.

[3] Edward H. Kennedy. Towards optimal doubly robust estimation of heterogeneous causal effects, 2023.

[4] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2009.

[5] Judea Pearl. The do-calculus revisited. *arXiv:1210.4852 [cs.AI]*, 2012.

[6] Nicholas J. Radcliffe and Patrick D. Surry. Real-world uplift modelling with significance-based uplift trees. White Paper TR-2011-1, Stochastic Solutions, 2011.

[7] Wendell R. Smith. Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, Vol. 21(No. 1):3–8, Jul., 1956.

[8] Peter J. Bickel Sören R. Künzel, Jasjeet S. Sekhon and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116.10, 2019.

[9] Tyler J. VanderWeele. Principles of confounder selection. *European Journal of Epidemiology*, 34(3):211–219, 2019.

[10] Robert Yi. Pylift: A fast python package for uplift modeling, 2018. Accessed: date.

[11] Yan Zhao, Xiao Fang, and David Simchi-Levi. Uplift modeling with multiple treatments and general response types. In *Proceedings of the 2017 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 2017.

About the Authors

Kaihua Ding is Director of Data Science and AI at AstraZeneca U.S., where he leads advanced data-science and machine-learning initiatives to boost sales-activity effectiveness, deploy trigger-based engagement solutions for field teams, and optimize speaker-program recommendations for healthcare providers. He holds a Ph.D. in Aerospace Engineering and Computational Science from the University of Michigan–Ann Arbor and serves as an elected board member of the National Science Foundation's ACCESS program.

Jingsong Cui is Senior Director of Commercial Data Science and AI at AstraZeneca U.S., where he designs and implements advanced analytics solutions for commercial operations in the biopharmaceutical industry. His expertise spans predictive modeling, customer segmentation, and analytics platforms that inform strategic decision-making across the organization.

Mohammad Soltani is Director of Data Science and AI at AstraZeneca U.S., specializing in commercial data science solutions for the Immunology and Respiratory (I&R) brands. He leads product strategy and evangelization efforts to ensure effective deployment and adoption of data-driven tools that optimize commercial performance.

Jing Jin is Director of Data Science and AI at AstraZeneca U.S., where she leads cross-portfolio AI and data science initiatives across the Cardiovascular, Renal and Metabolism (CVRM) brands. With over two decades of pharmaceutical industry experience and expertise in omni-channel orchestration, she champions new data-driven capabilities and fosters collaboration among marketing, medical affairs, and sales teams.

Can generative AI help to understand patient journeys more efficiently and effectively?

Gagan Bhatia, Roche/Genentech; Marco Giannitrapani, Roche; John Kaspar, Roche; Sharon Lee, Roche/Genentech; Suyash Mishra, Roche; Jessica Tashker, Ph.D. Roche

Abstract: Can generative AI help to understand patient journeys more efficiently and effectively? This is the question researched in this study by replicating patient journey mapping from diverse data sources, often done manually, by using generative AI tools. Two popular Large Language Models (LLMs) were tested - ChatGPT 4o and Claude Sonnet, with a Retrieval-Augmented Generation (RAG) model by uploading internally available research and external publications. Results were promising, with generative AI tools closely mirroring the human output when it comes to qualitative insights of the patient journey. For quantitative insights, however, the results were less accurate due to a variety of reasons including missing information in the source files, and inaccurate reading from complex data charts and graphs. Both LLMs - ChatGPT-4o and Claude Sonnet – had comparable performance. Study found that generative-AI tools enable increased efficiency by extracting the most critical information from available sources, allowing for more targeted follow-on research. However, given variable outputs for quantitative insights, the study recognized that these tools still require human review to ensure accuracy. It found that development of these tools and capabilities is a highly iterative process. As generative AI tools advance along with internal capabilities, there is hope to improve upon these limitations and integrate generative AI into the insights generation process. As companies create their own generative AI strategies and budget, this study provides learnings on expected value generation to inform their decision.

1. Introduction

Over the last two years, there has been an increasing interest in Generative AI, the newest wave of LLM technology, because of its promise to quickly and efficiently synthesize and generate insights from a variety of data sources. This article highlights the potential applications and limitations of generative-AI as related to insights generation and life sciences commercial strategy.[1] It spotlights a study related to patient journey mapping – the process of understanding the end-to-end experience of a potential patient for a medicine, from pre-diagnosis through ongoing treatment. This article discusses the design, results, and takeaways from a control experiment aimed at evaluating if gen AI can replicate and/or improve upon the existing patient journey mapping process.

2. Methods / Approach

Overview:

The current patient journey mapping process is time intensive and expensive. Typically, insights leaders pull from a number of sources, including: primary market research studies, syndicated reports, field insights, medical publications, and internal subject matter expertise. Each component part could take months to compile, and often are quite expensive.

The study's aspiration was to improve the patient journey mapping process by leveraging generative AI. To test its potential, a "control" experiment was designed to mirror the most critical questions associated with the patient journey mapping process. The experiment was completed using two different generative AI LLMs: GPT-4o and Claude Sonnet. A Retrieval-

Augmented Generation (RAG)[2] approach was used, where it only references the trusted data sources provided for accuracy and reliability of response, however, study also tested a scenario where no data was provided and only the foundational-trained model was used.

Objectives for the experiment were as stated below:

Central Questions:

- Can AI replicate an existing patient journey map (control experiment) by accurately answering a set of 8 targeted prompts?
- Can AI improve the patient journey mapping process by improving efficiency, generating new insights, or otherwise?

Objectives:

1. Understand if a prompt-based generative AI tool based on LLMs (i.e., GPT-4o or Claude Sonnet) can replicate and/or improve the patient journey mapping process
2. Assess the value / quality of outputs for GPT-4o relative to Claude Sonnet
3. Identify pain points and opportunities to further enhance existing tools, with the goal of leveraging for future patient journey development

Design:

The control experiment was designed via a 3-step process (Exhibit 1).

A) First, the study team reviewed a previously completed patient journey map to formulate eight central questions that mirror the most important findings from the mapping process. These eight questions served as prompts to be asked to a gen-AI tool. Questions were split between those qualitative vs quantitative in nature. "Benchmark" responses were recorded using a completed patient journey map for a particular disease area with what was deemed the "correct" response. Example questions include:

- Qualitative: How can we define a disease X patient in terms of time on disease, symptoms, therapies, support needed, and challenges they face?
- Quantitative: What are the comorbidities that disease X patients have? What percentage of patients have each of the comorbidities?

B) Second, data from all sources was gathered from the patient journey study, including primary market research, syndicated research reports, medical journal publications, and internal field insights. These files were uploaded to each gen-AI tool corresponding to the LLMs, to provide additional context when answering the 8 prompts.

To extract and chunk the data, first the pages were tagged with document name and page number, then chunks were created using recursive character text splitter and further these chunks were tagged with their chunk number, page number, and document name. A small but useful improvement was maintaining the original formatting of lines and paragraphs during extraction.

C) Third, all 8 prompts were tested across both tools and three different scenarios – 48 total permutations. The scenarios tested include:
1) **All data sources** including primary market research (containing the existing patient journey), syndicated research, medical

publications, and field insights; 2) **All data sources excluding primary market research** containing the previous patient journey mapping; 3) **No data sources** where only the foundational model[3] was used. These scenarios demonstrate how these gen-AI tools process provide varying degrees of support. Scenario 2 is the most likely scenario that mirrors the information that would be available at the time a draft patient journey would be created, prior to completion of primary market research. Ideally, it is in this scenario where generative AI tools could provide the most value in the long term.

To minimize contradictions from input sources, various techniques were experimented with, e.g. tuning chunk overlap, chunk length, and top_k (number of results retrieved) to optimize retrieval and prevent contradictory inputs to the model. Other experiments are ongoing to test if it's preferable to review contradictory results by the user and feedback and preference be integrated into the tool.

Few-shot prompting was employed in the model where examples of format of outcomes were provided. A query bifurcation function was implemented, that took care of queries with multiple intents. It bifurcated queries into smaller queries, generating responses to those queries and then collating the final query.

All three scenarios were tested and scored using tools enabled by GPT-4o and Claude Sonnet. Both generative AI platforms allow users to upload PDFs and ask targeted prompts.

Evaluation:

Each prompt was graded using a simple three stage scoring system. Other statistical and score based measures were tested and deemed biased and inappropriate for this evaluation. See details on scoring below:

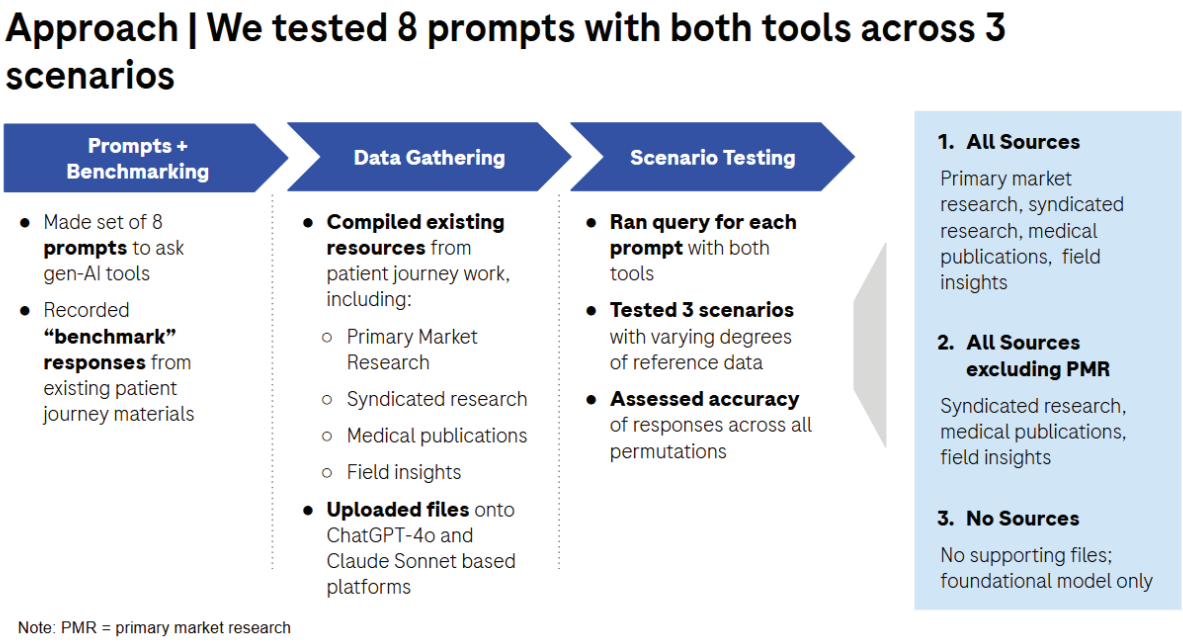
A. **Inaccurate Responses (Red)**
   Did not match values / sentiment from existing patient journey

B. **Incomplete Responses (Grey)**
   Did not fully address question (i.e. left off important details if qualitative; some values not provided and/or inaccurate if quantitative)

C. **Comparable Quality Output (Green)**
   Response is accurate and mirrors output from existing patient journey

This scoring approach favors the scenario where a gen-AI tool recognizes its limits and does not attempt to provide a component of an answer that may be uncertain (i.e., due to a lack of data or knowledge of where to access such data), rather than the scenario where an improper value is provided. Furthermore, the study recognizes that in the early stages of using gen AI tools, manual review will be required for all outputs to ensure accuracy / traceability of response provided.

Exhibit 1:



3. Results

Findings from this experiment ultimately demonstrate both the promise that Generative AI holds in supporting insights generation in the near future, as well as the degree of work that still needs to be done to drive full-scale adoption of these tools.

As mentioned, questions were categorized as qualitative or quantitative – results of this experiment fared significantly better for qualitative prompts vs quantitative ones. Agnostic to which gen-AI LLM is used (i.e., GPT-4o or Claude Sonnet), qualitative prompts demonstrated a comparable output (green) for 20/24 (83%) prompts, while quantitative prompts achieved this result just 1/24 (4%) times. Likewise, inaccurate (red) outputs were never generated for qualitative prompts in 24 attempts but 50% of the time (12/24) for quantitative prompts (Exhibit 2, Exhibit 3).

The differences in these results are profound. Evidently, both generative AI LLMs are able to better assess qualitative prompts and were unable to accurately reference quantitative data. The study team continues to improve the ability of these tools to accurately pull from charts, graphs, tables, and other standard ways to represent data – this is a work in progress. Some of the ways this is being addressed are by a combination of OCR vision models and improving content enrichment pipelines (contextualisation, transcription, tagging and captioning).

When comparing the outputs from GPT-4o and Claude Sonnet, results are comparable. GPT-4o results were polarizing with 12/12 qualitative prompts earning comparable (green) outputs, and 0/12 quantitative prompts. The Claude Sonnet earned 8/12 and 1/12 comparable

(green) outputs, respectively. GPT-4o also showed a slightly higher likelihood of inaccurate responses, earning Inaccurate (Red) marks for 7/24 prompts vs 5/24 prompts using Claude Sonnet. Results across the two LLMs are too similar to conclude that one LLM should be preferred over the other and additional testing will be required.

Exhibit 2:

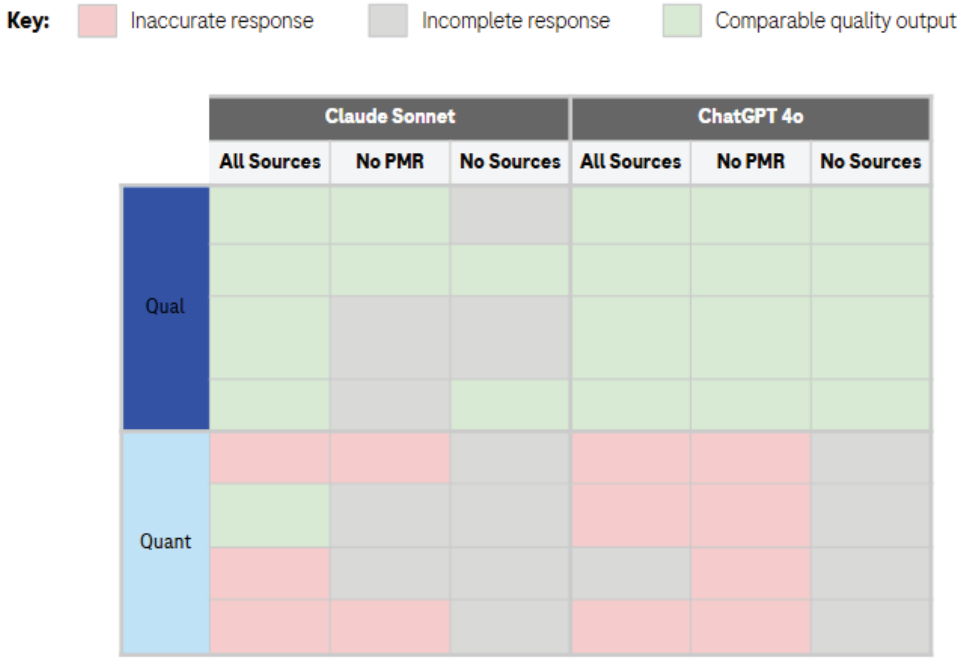
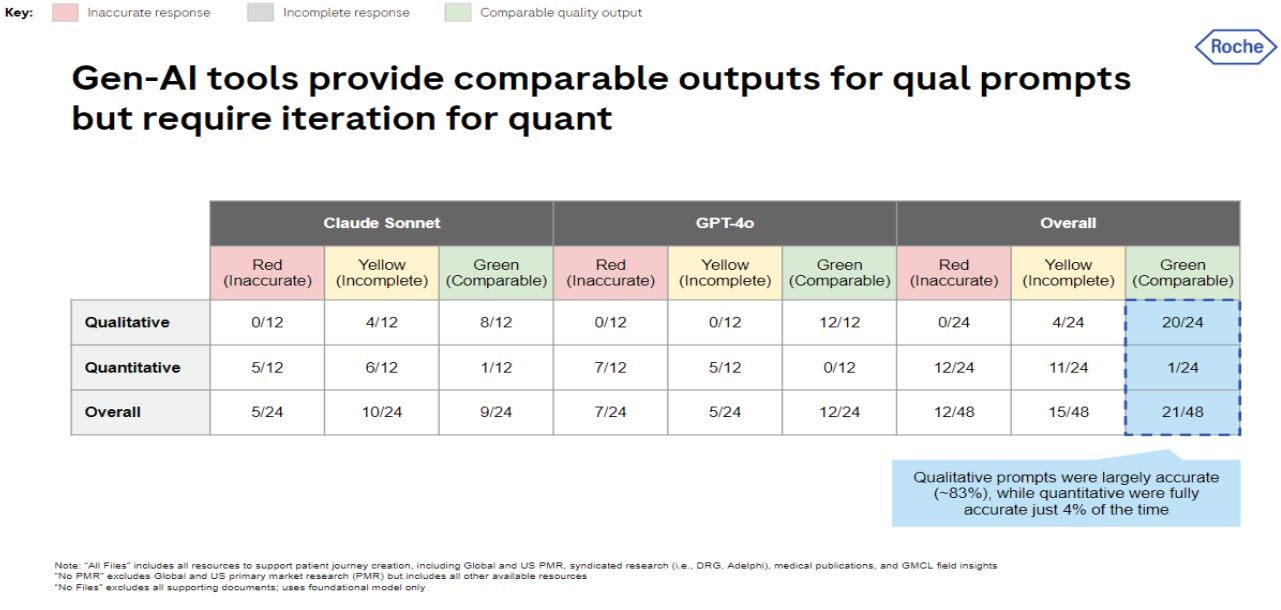


Exhibit 3:

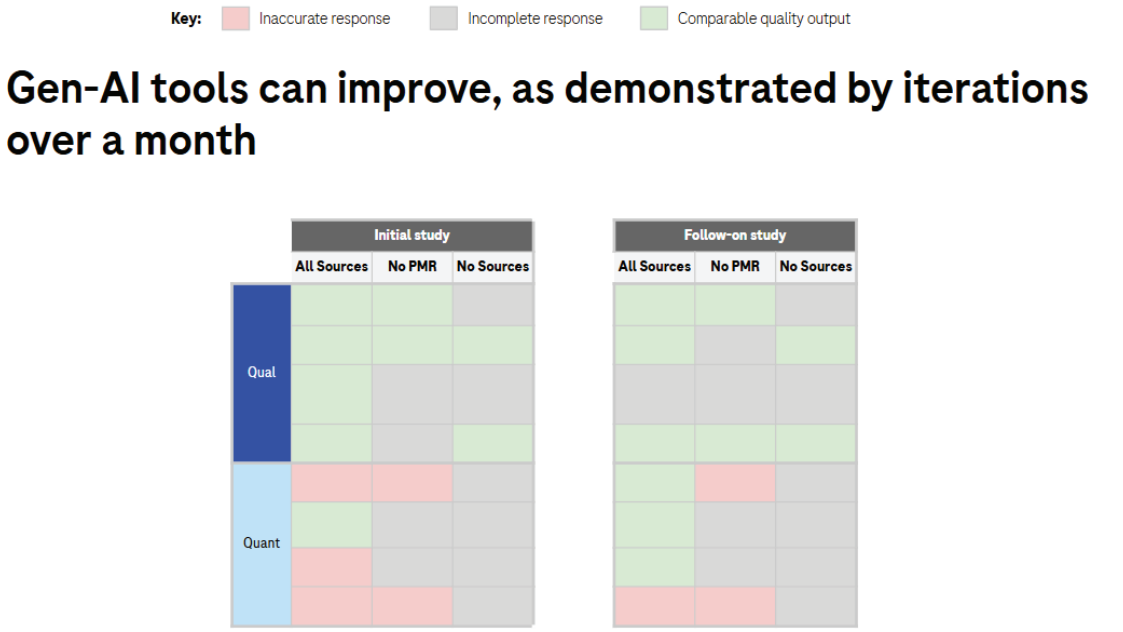


After these initial results, the existing difficulties faced with these generative-AI LLMs were documented– both from an accuracy and user interface perspective. Over the course of ~4 weeks, the analytics team worked to refine the underlying model (including prompt engineering) for the Claude Sonnet based tool. With these adjustments completed, a follow-on study was conducted with just the Claude Sonnet based tool, across the same three scenarios previously outlined.

Results from this follow-on trial were promising (Exhibit 4). The Claude Sonnet tool demonstrated considerable improvement in its ability to accurately interpret charts / graphs and answer quantitative prompts. However, there was additional variance in the results for more qualitative questions.

These findings show that significant progress can be made in the development of generative-AI tools in just one month. There are still imperfections with these tools to work through, but the promise of these technologies and their ability to continue to grow in utility is evident.

Exhibit 4:



4. Discussion / Conclusions

While this study is just one control experiment, there are learnings that can be used to embark on a long-term journey to incorporate generative AI technology into many other use cases. Following are the key key takeaways:

Qualitative responses work better than quantitative: Since current LLMs are very powerful at reading and generating text based data, they perform very well at it. In this study, the best model produced 100% comparable results for qualitative responses, while the best model produced 25% comparable results for quantitative responses, when compared with the benchmark. This increases confidence that these tools can be used with high reliability where qualitative inputs and outputs are involved. However, when it comes to quantitative data, accuracy is lower due to complexity of charts and tables or unavailability of reliable quantitative data in input sources. Further enhancements are required to make these tools more reliable for quantitative analysis.

Performance is similar across LLMs: Both LLMs were very close in their performance and there was no meaningful difference. This shows the suitability of both LLMs in using a RAG based approach. However, as the goal is to also improve accuracy of responses to quantitative outputs, one must look at future evolution and capabilities of these LLMs.

Performance measurement: Performance measurement was subjective but it worked reasonably well in assessing the results. While there are multiple ways to look at performance,[4] the study kept the performance metrics simple and intuitive to avoid machine or human biases.

Iterations are powerful: One key learning from this experiment is that the performance of the gen AI tool improves with iterations on data, prompts and other user interface features. Most of the study iterations were spent on prompt engineering. Hence, one should plan multiple iterations and not get discouraged if initial results show low performance.

Some examples of prompt engineering employed include the following protocol:

- While stating examples to get a specific response type, explained all the possible examples in the system prompt for better results
- Context that is retrieved and passed in the model was well formatted
- Refrained from repeating instructions in the system prompt
- Instead of negative words like can't, don't, can not, mentioned "refrain" if something needed to be avoided
- Tables were passed separately in the system prompt
- Gave filtered information based on tags during the retrieval part for both text and tables, which resulted in better responses
- See below for prompt wording which was critical in improving responses

Prompt wording is critical: Based on this study, both GPT-4o and Claude Sonnet platforms provided wordy responses when prompts were open-ended, requiring a trial-and-error process to determine the proper wording for the 8 prompts. It showed the importance of asking questions as explicitly as possible, providing clear topics for the gen-AI

platform to address. For example, a prompt that asks, “What challenges do disease X patients often face?” is likely to receive a generic response. A more targeted prompt might be “In ~5 bullets, can you summarize the challenges patients with severe disease X face? Include symptoms, relevant therapies, and additional support needed in your summary.” The addition of these specific details in question wording are critical to ensuring higher quality outputs. Some best practices in generating prompt are listed here.[5]

Cross functional collaboration: User feedback from insights and business users is very important, as it can provide input on where the performance is lacking and to identify areas of improvement for the iterations. Close collaboration between data science, insights and business functions is key to a successful experiment and future implementation.

Human review is paramount: While LLMs were able to retrieve and cite references for information sources; it still required human review to validate and trust. This is especially needed when quantitative information is being cited which can sometimes be inaccurately pulled. In addition, if the model is asked to produce any derived information, like averages, the calculations need to be checked. This however means, that in the short term, any efficiency gains may be partially offset by time required to review and validate the results.

User interface is a key factor for ease of use and future adoption: A final learning is on the importance of the user interface used for applying the gen AI technology. Various user interface features like ability to upload data, provide guiding prompts, generating well formatted outputs, ability to copy results in other documents, and easy review of references, are very important in ensuring a smooth user experience and ultimately, the adoption of the technology. For example, the study initially faced some processing issues, which led to higher compute time for responses, and it became a significant pain point until it was resolved. Developers of gen-AI based tools must keep the user interface design an important part of the solution for enabling adoption.

This study showed several areas where generative AI technology performs very well, i.e. in generating qualitative responses and providing efficiency in generating insights from multiple sources. There are still limitations, e.g. in generating accurate quantitative insights and need for human review and validation. However, as seen by improvement through iterations on data, prompts and user interface, and continuous improvements to available foundational LLM models, there is promise that gen AI based tools can significantly add value to the insight generation process. The study provides several learnings that can be built upon to make gen AI solutions more effective.

References:

1. Asli Aksu, Rukhshana Motiwala, Sherin Ijaz, Nicholas Mills, Ashley van Heteren, Oscar Viyuela Garcia, *Early adoption of generative AI in commercial life sciences*, Mckinsey & Company, May 6 2024. Available from: https://www.mckinsey.com/industries/life-sciences/our-insights/early-adoption-of-generative-ai-in-commercial-life-sciences

2. *What is RAG (Retrieval-Augmented Generation)?*, AWS. Available from: https://aws.amazon.com/what-is/retrieval-augmented-generation/

3. *What are Foundation Models?*, AWS. Available from: https://aws.amazon.com/what-is/foundation-models/

4. Giovanni Maria Iannantuono, Dara Bracken-Clarke, Fatima Karzai, Hyoyoung Choo-Wosoba, James L Gulley, Charalampos S Floudas, *Comparison of Large Language Models in Answering Immuno-Oncology Questions: A Cross-Sectional Study*, Oncologist,February 3 2024. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC11067804/

5. *Prompt Engineering - Six strategies for getting better results*, OpenAI. Available from: https://platform.openai.com/docs/guides/prompt-engineering/

Acknowledgment

The authors would like to thank Samik Adhikary (Head of Products and Services for Advanced Analytics at Roche) and Jean-Paul Abbuehl (Data Science Strategy Lead for Advanced Analytics at Roche) for their input and support of this work.

About the Authors

Gagan Bhatia is a Global Catalyst for Advanced Analytics at Roche/Genentech. He has vast experience in Pharma data and analytics for functions spanning sales, marketing, access and portfolio strategy, both as a consultant and industry leader. He has worked with several large pharma and small biotechs to create data driven business strategies and operational implementations. Gagan is currently leading multiple AI initiatives at Roche to revolutionize the way data and insights are created and consumed.

Marco Giannitrapani is the VP of Commercial Data, Analytics and Marketing Technology at Roche. He has over 20 years of experience in enabling companies to leverage data, analytics and technologies to drive business outcomes. Marco has his largest experience in the Pharmaceutical sector, however he also worked in the Financial Services, Oil & Gas, and Education. Marco holds a PhD in Statistics from the University of Glasgow (UK), an MBA from IESE Business School (Spain), and an MSc in Statistics and Economics from University of Palermo (Italy).

John (Jack) Kaspar is an intern in Roche's Ecosystem Insights group, where he drives market and customer centric insights to support Roche's portfolio. He's currently completing his M.B.A. at the University of Chicago Booth School of Business, earning concentrations in Healthcare, Finance, and Entrepreneurship. Jack previously spent 4 years in healthcare consulting and studied biology at the University of Notre Dame.

Sharon Lee, Ph.D., is the Immunology Integrated Insights Leader at Roche/Genentech, with over 10 years of experience spanning scientific and commercial excellence. Her background includes roles as a drug discovery chemist, marketing leader, and insights leader. Currently, Sharon is passionate about bridging her deep scientific, commercial, and insights expertise to test and improve new applications like Gen AI to enable better business decisions.

Suyash Mishra serves as Lead Data Scientist within Roche's Global Pharma Strategy Group. He leverages deep expertise in NLP and Graph Theory to pioneer AI solutions across the content value chain. He leads the development of innovative AI solutions that optimize the entire content value chain from creation to analytics. Suyash's current focus is on using Generative AI to build digital twins that accelerate the creation (net new/ derivative) and adaptation of highly personalized content.

Jessica Tashker, Ph.D., is the Global Head of Ecosystem Insights at Roche Pharmaceuticals. With 15+ years of experience in the Insights field, she has depth of expertise in primary market research, forecasting, and secondary data analytics. Jessica's current passion includes seeing how new applications of Gen AI can empower us as Insights professionals to have an even greater impact on the business and the patients we serve.

Multimodal Multi-Agent Solution for Sales Representatives

Shihan He[1] and Anubhav Srivastava[2]

[1]Machine Learning Engineer at AI CoE, Novo Nordisk Inc. and [2]Associate Director at AI CoE, Novo Nordisk Inc.

Abstract

The relationship between pharmaceutical sales representatives and Healthcare Professionals (HCPs) is undergoing a significant transformation, moving from a historically transactional focus to an era characterized by the necessity of genuine, trust-based partnerships. This evolution is driven by a mutual understanding of the ultimate goal: improving patient outcomes. Establishing and nurturing these authentic connections is not merely beneficial but critical for the future of healthcare collaborations. In fact, a substantial majority of HCPs have indicated a significantly higher likelihood of engaging with a sales representative if the interaction mirrors the quality of their best professional relationships. Furthermore, positive engagements can foster broader company affinity, increasing the propensity of HCPs to open emails and pay attention to a company's messages.

Effective engagement between pharmaceutical companies and HCPs is paramount, serving as the vital link through which HCPs receive the latest information on drugs, treatment protocols, and clinical pipelines, while also providing a platform for them to ask questions, voice concerns, and offer valuable feedback. Pharmaceutical sales representatives are at the forefront of this engagement, playing a multifaceted role that extends beyond simply promoting and selling drugs. Their responsibilities include providing comprehensive and accurate education about pharmaceutical products, building and maintaining strong relationships with HCPs, identifying and pursuing new sales opportunities, and ensuring adherence to ethical standards and regulatory guidelines. Given the increasing complexity of the pharmaceutical industry and its stringent regulatory environment, the role of certified pharmaceutical representatives has become even more significant. This complexity underscores the necessity for sales representatives to be well-versed in scientific and medical aspects of products, fully understand regulatory compliance, and possess strong communication and relationship-building skills.

To address these challenges, this paper proposes a multimodal multi-agent system designed to empower sales representatives with advanced tools and real-time insights. The system leverages the capabilities of three distinct agents: a Natural Language Processing (NLP) agent to analyze HCP interactions and extract key information, a machine learning agent to predict HCP behavior and preferences, and a reporting agent to streamline reporting tasks and provide actionable summaries.

By facilitating efficient knowledge sharing and collaboration among these agents, the system enables sales representatives to gain a deeper understanding of HCP needs and tailor their engagement strategies accordingly. The system's knowledge base is continuously updated with the latest data and insights, ensuring that sales representatives have access to the most relevant information for informed decision-making.

This paper explores potential use cases and applications of the system, demonstrating its versatility in various scenarios such as product launches, new HCP targeting, and ongoing engagement efforts. The paper also discusses methods for evaluating and validating the system's performance, ensuring its effectiveness and accuracy in real-world settings.

Finally, the paper discusses potential benefits and limitations of the proposed solution, outlining future research directions and potential enhancements to further optimize its capabilities. The multimodal multi-agent system presented in this paper has the potential to significantly impact pharmaceutical sales and HCP engagement by empowering sales representatives with advanced tools and real-time insights for informed decision-making and tailored engagement strategies.

Key words: GenAI Solutions, Multimodal, Multi-Agent, RAG System

1. Introduction

Content Enablement for Effective HCP Interactions: Challenges and the Critical Role of Sales Representatives

The pharmaceutical industry operates within a dynamic and highly regulated environment, where the engagement between sales representatives and Healthcare Professionals (HCPs) is paramount. These interactions are crucial for disseminating vital information about new treatments, medical devices, and therapeutic advancements. However, the sheer volume and complexity of medical knowledge, coupled with the increasing sophistication of HCPs, necessitate a paradigm shift in how sales teams are equipped and supported. Traditional digital tools, while offering a degree of access to information, often fall short of providing the nuanced knowledge and strategic guidance required for truly effective engagements.

The current landscape demands that sales representatives not only possess a deep understanding of the products they represent but also the broader clinical context, relevant research, and the specific needs and preferences of individual HCPs. Merely providing access to a vast repository of documents, presentations, and marketing materials is insufficient. Sales teams require intelligent tools that can surface the right information at the precise moment of need and, crucially, guide them on how to leverage this information strategically during their preparation for HCP interactions. The healthcare sector's stringent regulatory framework further underscores the importance of accuracy and compliance in all communications with HCPs, making robust content enablement solutions an indispensable asset. The increasing volume and complexity of medical information, combined with the imperative for strategic application in HCP interactions, reveals a significant gap that conventional tools struggle to bridge.

This necessitates the adoption of intelligent, context-aware solutions, such as a multi-modal, multi-agent Retrieval-Augmented Generation (RAG) system. This system represents a fundamental advancement in internal content enablement for pharmaceutical sales representatives. It moves beyond simply providing access to information; it actively assists representatives in finding, understanding, and utilizing the most relevant knowledge for each specific situation.

Recognizing that sales enablement materials span diverse formats, our proposed RAG system features a multi-modal architecture. It's designed to ingest, process, and integrate information from various sources common in sales environments—including text documents, slide presentations, images, training videos, and audio recordings. This holistic approach mirrors how sales representatives naturally encounter and utilize information, aiming to improve their knowledge retention and practical application. Furthermore, the system employs a hierarchical multi-agent architecture, coordinated by a Task Maestro Agent that utilizes chain-of-thought reasoning to manage complex requests. This central agent delegates specific sub-tasks to a team of 'Synergistic Content Agents', each with a distinct specialization, such as the Resource Locator for finding content, the Market Insight Agent for relevant data, the Tailoring Agent for personalization, the Slides Crafter for presentation elements, and the Content Optimizer for refining output. This collaborative framework ensures that sales representatives receive not just accurate information, but contextually relevant content optimized and formatted for their immediate needs.

In this article, we detail the design and implementation of a multi-modal, multi-agent RAG system and demonstrate its potential to improve the effectiveness and efficiency of pharmaceutical sales representatives.

2. Preliminary

2.1 Understanding Retrieval-Augmented Generation (RAG)

At its core, Retrieval-Augmented Generation (RAG) represents an advanced AI framework designed to enhance the capabilities of Large Language Models (LLMs) by grounding their responses in external, verifiable knowledge sources. This process involves two primary stages: first, retrieving relevant information from a designated knowledge base, and second, utilizing this retrieved information to augment the LLM's content generation process. By incorporating external data, RAG helps to mitigate inherent limitations of LLMs, such as the potential for generating outdated or factually incorrect information, often referred to as hallucinations. This foundational mechanism of RAG is crucial for providing sales representatives with access to and the ability to utilize the extensive amounts of digital content relevant to their interactions with HCPs. It ensures that the information delivered is firmly rooted in factual data, thereby significantly reducing the risk of inaccuracies, a paramount concern within the healthcare domain.

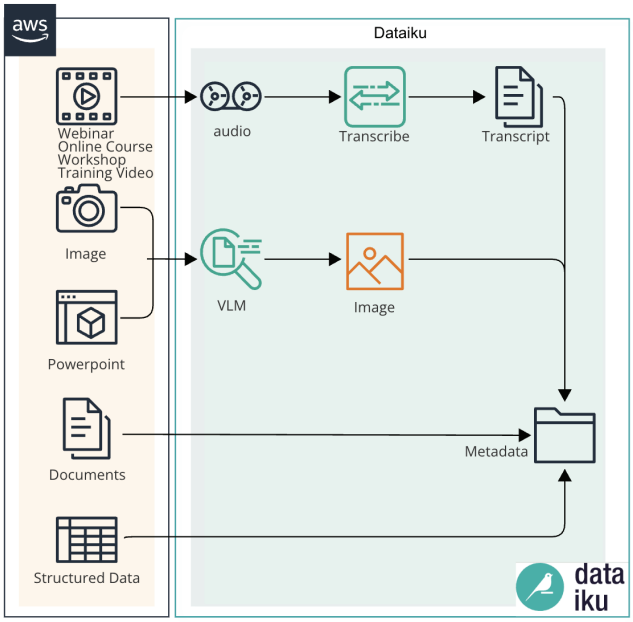
2.2 The Power of Multi-Modality in HCP Interactions

While traditional RAG primarily focuses on textual data, multi-modal RAG extends this framework to encompass a wider array of data formats, moving beyond text to include images, tables, images and potentially even audio and video content. In the context of HCP engagement, the information landscape is rich and diverse, often incorporating conference pictures, clinical trial results presented in tables, and educational videos explaining complex medical procedures. A multi-modal solution possesses the capability to process and seamlessly integrate information derived from these disparate sources, thereby offering a more comprehensive and nuanced understanding of the subject matter. These diverse forms of content often enrich interactions with HCPs, and a multi-modal RAG solution can effectively leverage them. This provides sales representatives with a more holistic understanding and more impactful communication tools, enabling them to engage with HCPs using the full spectrum of relevant information.

2.3 Multi-Agent Systems for Enhanced Guidance

To further enhance the effectiveness of RAG for sales enablement, the integration of multi-agent systems offers a significant advantage. Multi-agent systems involve the collaborative efforts of multiple autonomous AI agents working in concert to achieve a shared objective. Within the realm of sales enablement, these agents can be specialized to handle specific tasks, such as the efficient retrieval of knowledge, the provision of strategic analysis based on that knowledge, and the generation of actionable content tailored for HCP interactions. To ensure a cohesive and efficient operation, a central orchestrator agent can be implemented to manage the overall workflow and facilitate seamless collaboration among these specialized agents. This multi-agent approach directly addresses the user's need for "strategy." By having agents specifically designed to analyze retrieved knowledge and understand the context of HCP interactions, the system can provide strategic recommendations that go beyond simply delivering information.

3. Methodology



3.1 Large Language Models (LLMs) and Vision Language Models (VLMs)

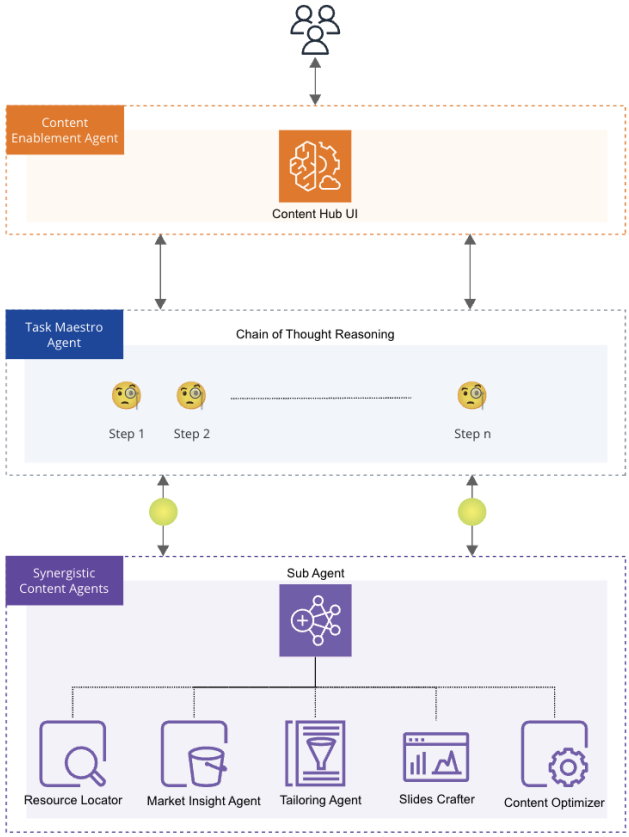
The foundational intelligence of a multi-agent multi-modal RAG solution relies heavily on the capabilities of Large Language Models (LLMs) and Vision Language Models (VLMs). LLMs serve as the primary generative engine, adept at processing text-based information and generating coherent and contextually relevant responses. The model we use is **azure-openai-gpt4o**. Complementing LLMs are VLMs, which extend these capabilities to the visual domain, enabling the system to understand and process information contained within images and potentially videos. Examples of VLMs include in the solution is **Anthropic Claude 3.5 Sonnet**. The careful selection of appropriate LLMs and VLMs is paramount for the system's ability to effectively understand and generate content across the diverse modalities encountered in HCP interactions.

3.2 Embedding Models for Text, Images, and Tables

To facilitate efficient information retrieval across different data types, embedding models play a crucial role. These models convert text, images, and tables into high-dimensional vector representations, allowing the system to perform semantic similarity searches and retrieve relevant information based on meaning rather than just keywords. This solution includes the **azure-openai-embed model** for text and the **Amazon Titan image embedding model** for images. It is often necessary to employ different embedding models tailored to the specific characteristics of each modality to accurately capture their unique semantic features. The quality of these embedding models is critical for the Knowledge Retrieval Agent to effectively locate pertinent information from the organization's digital tools, regardless of the content format.

3.3 Vector Databases for Efficient Information Retrieval

The efficient storage and retrieval of the vector embeddings generated by the embedding models are handled by specialized databases known as vector databases. These databases are specifically designed to store and query high-dimensional vector data, enabling fast and accurate similarity searches. Examples of vector databases include **FAISS** and **ChromaDB**. These databases are essential for rapidly identifying and retrieving the most relevant content in response to a sales representative's query during time-sensitive interactions with HCPs. A scalable and performant vector database is therefore a fundamental requirement for a RAG solution aiming to provide real-time guidance.



3.4. Multi-Agent System for Content Enablement

The diagram illustrates a multi-agent system designed for Content Enablement, structured in distinct layers to process user requests initiated via the **Content Hub UI**:

- **Orchestration Layer - Task Maestro Agent:** This central agent acts as the primary orchestrator. It receives tasks from the UI, interprets the user's needs, and manages the overall workflow. It utilizes **Chain of Thought Reasoning** to break down complex requests into sequential steps (Step 1, Step 2... Step n).
- **Execution Layer - Synergistic Content Agents:** This layer consists of multiple specialized agents working collaboratively to fulfill the task steps defined by the **Task Maestro Agent**. **Sub Agent** appears to coordinate the activities within the **Synergistic Content Agents** layer, receiving instructions from the **Task Maestro Agent** and delegating to the appropriate specialist agent.
- **Specialized Agents:**
  - **Resource Locator:**Responsible for finding and retrieving relevant information such as clinical study results, approved prescribing information, formulary details, treatment guidelines, and existing compliant content assets (e.g., detail aids, slide decks) pertinent to the HCP's specialty or query.

- **Market Insight Agent:** Tasked with gathering and analyzing market-specific data including therapeutic area trends, competitor activities, payer policies, and insights specific to the HCP (e.g., specialty, prescribing behavior, past interactions, identified clinical needs, institutional affiliations).
- **Tailoring Agent:** Focuses on customizing or adapting retrieved information and core content assets to the specific context of the HCP interaction, considering the HCP's specialty, patient population focus, prior engagement history, and anticipated clinical questions or objections.
- **Slide Crafter:**Specializes in generating relevant presentation slides or visual content incorporating clinical data visualizations, mechanism of action (MoA) diagrams, patient journey illustrations, or formulary comparison visuals, designed for effective and compliant HCP engagement.
- **Content Optimizer:**Works on refining and improving the generated or tailored content, ensuring medical accuracy, adherence to regulatory and compliance guidelines (including fair balance and appropriate referencing), brand consistency, and overall quality suitable for ethical and effective HCP communication.

3.5. Agent Orchestration Framework

The development and management of the interactions between these multiple agents can be significantly simplified through the use of agent orchestration frameworks such as **Dataiku Advanced LLM Mesh** feature. This frameworks provide developers with the necessary tools to define agent workflows, establish communication protocols between agents, and manage the overall decision-making processes within the system. Utilizing a robust agent orchestration framework streamlines the development process and allows for easier scaling and customization of the multi-agent system over time.

This study presents a multimodal multi-agent solution designed to enhance the efficiency of pharmaceutical sales representatives in their interactions with healthcare professionals (HCPs). The system leverages text, voice, and visual data inputs to capture and analyze interactions in real-time, providing actionable insights and streamlined reporting.

4. Outcome

Transforming Sales Interactions: Providing Knowledge and Strategy Through Content

4.1 Enhanced Access to Comprehensive and Up-to-Date Knowledge

A key advantage of a multi-agent multi-modal RAG solution is its ability to integrate seamlessly with the various existing digital tools utilized by sales representatives, including Customer Relationship Management (CRM) systems, centralized document repositories, and marketing automation platforms. This integration allows sales representatives to pose natural language questions and receive relevant information aggregated from across these diverse systems in a unified and easily digestible format. Furthermore, the inherent RAG component of the solution ensures that the information retrieved is the most current and accurate data available within the connected systems. By providing a single, intelligent point of access to knowledge from multiple sources, the solution significantly reduces the time sales representatives spend searching for critical information, thereby ensuring they have the most up-to-date content readily available for their interactions with HCPs.

4.2. Contextualized Information Delivery for Personalized Interactions

The multi-modal capabilities of the RAG solution enable it to understand the specific context of each HCP interaction. This includes considering factors such as the HCP's medical specialty, records of past interactions, and any previously expressed interests or preferences. The information retrieved by the system can then be precisely tailored to this specific context, ensuring that the sales representative delivers the most relevant and personalized content to the HCP. This level of personalization is crucial for fostering stronger relationships with HCPs and ensuring that the information shared is both timely and pertinent to their needs.

4.3. Strategic Guidance Integrated into the Workflow

Beyond simply providing access to knowledge, the strategy recommendation agent within the multi-agent system plays a vital role in enhancing the effectiveness of sales interactions. This agent can analyze the available information, taking into account the specific context of the HCP interaction, and suggest the most effective approach for the sales representative to adopt. This includes recommending key talking points to emphasize, anticipating potential questions the HCP might ask, and providing guidance on how to proactively address potential objections or concerns 16. This strategic guidance can be delivered in real-time during an interaction or as part of pre-call planning, empowering sales representatives to engage with HCPs in a more informed and strategic manner, ultimately leading to more productive and successful engagements.

5. Practical Applications and Use Cases for Sales Reps Interacting with HCPs

The implementation of a multi-agent multi-modal RAG solution can significantly enhance various aspects of a sales representative's workflow when interacting with HCPs.

5.1. Pre-Call Planning and Preparation

Before engaging with an HCP, sales representatives can utilize the system to quickly gather comprehensive background information. This includes details about the HCP's research interests, recent publications, and a history of past interactions with the company. The system can also provide easy access to relevant product information, including clinical trial data and approved marketing materials in various formats such as text, images, and tables. Furthermore, the strategy recommendation agent can analyze this information and suggest key areas of focus for the upcoming meeting, ensuring the sales representative is well-prepared and can tailor their message effectively. This enhanced pre-call planning leads to more focused and productive meetings.

5.2. Real-Time Support During HCP Interactions

During a meeting with an HCP, the sales representative can leverage the system to provide immediate and accurate responses to questions that may arise. If an HCP inquires about specific details or data, the sales rep can use natural language to query the system and retrieve the most up-to-date information in real-time. The multi-modal capabilities allow the representative to seamlessly share relevant visuals, such as mechanism of action diagrams or data tables summarizing clinical trial results, to support their answers and enhance the clarity of their communication. The system can also offer real-time guidance on how to best respond to specific questions or concerns raised by the HCP, enabling the sales representative to handle inquiries with confidence and accuracy.

5.3. Post-Call Follow-Up and Action Items

Following a meeting with an HCP, the system can assist the sales representative in summarizing the key discussion points and identifying any necessary follow-up actions. Based on the conversation, the system can suggest relevant resources, such as specific research papers or detailed product brochures, to share with the HCP. Additionally, the system can help the sales representative efficiently update the CRM system with crucial information gathered during the interaction and assist in planning for future engagements with the same HCP. This streamlined approach to post-call activities ensures that sales representatives can effectively follow up and maintain positive momentum in their relationships with HCPs.

6. Benefits of Implementing a Multi-Agent Multi-Modal RAG Solution

The adoption of a multi-agent multi-modal RAG solution offers a multitude of benefits for pharmaceutical and medical device companies seeking to enhance their sales interactions with HCPs.

6.1. Enhanced Sales Productivity and Effectiveness

By providing sales representatives with readily accessible and contextually relevant knowledge and strategic guidance, this solution empowers them to have more productive and successful interactions with HCPs 4. They can respond to inquiries with greater accuracy and provide more pertinent information, thereby bolstering their credibility and fostering trust with the healthcare professionals they engage with. This improvement in sales productivity directly contributes to better business outcomes, including increased sales figures and the cultivation of stronger relationships with key stakeholders in the healthcare community.

6.2. Improved HCP Engagement and Satisfaction

HCPs benefit from receiving timely, accurate, and personalized information tailored to their specific needs and interests, leading to a more positive and valuable interaction experience 4. This enhanced engagement and satisfaction can translate into stronger, more enduring relationships between the company and the HCPs it serves. Satisfied HCPs are more likely to be receptive to the company's products and services, which is crucial for long-term success in the competitive healthcare market.

6.3. Increased Efficiency and Reduced Costs

The implementation of this solution leads to increased efficiency within the sales team. Sales representatives spend less time searching for necessary information and can dedicate more of their time to direct engagement with HCPs 4. Furthermore, the system can automate certain routine tasks, such as the retrieval of information and the initial stages of post-call follow-up, thereby freeing up valuable time for sales representatives to focus on more strategic activities. This improved efficiency and automation can result in significant cost savings for the organization over time.

6.4. Enhanced Compliance and Reduced Risk

In the highly regulated healthcare industry, ensuring compliance with relevant regulations and company policies is paramount. A multi-agent multi-modal RAG solution can be specifically designed to ensure that all information shared with HCPs adheres to these stringent requirements 6. This capability significantly reduces the risk of sales representatives inadvertently sharing inaccurate or unapproved information, thereby safeguarding the company from potential regulatory issues and reputational damage.

6.5. Scalability and Adaptability

The architecture of a multi-agent multi-modal RAG solution is inherently scalable, allowing it to accommodate a growing volume of digital content and an expanding base of users as the organization's needs evolve 16. The modular design of the system also allows for its adaptation to incorporate new data sources and to evolve in response to changing market dynamics and emerging information. This scalability and adaptability ensure the long-term value and sustained relevance of the solution to the organization's strategic objectives.

7. Key Considerations and Challenges for Successful Implementation

While the potential benefits of a multi-agent multi-modal RAG solution are substantial, successful implementation requires careful consideration of several key factors and potential challenges.

7.1. Data Integration and Quality

A critical prerequisite for the success of the RAG solution is the effective integration with the organization's existing digital tools and the assurance of high data quality and consistency. The data from these various sources needs to be properly indexed and readily accessible to ensure efficient and accurate retrieval by the system. The overall effectiveness of the RAG solution is directly contingent upon the quality and accessibility of the underlying data. Therefore, meticulous planning and execution of the data integration process are essential.

7.2. Agent Orchestration Complexity

Designing and managing the intricate interactions between multiple autonomous agents within the system can present a significant challenge and necessitates careful planning and architectural design. Ensuring seamless communication and effective collaboration among these specialized agents is paramount for achieving optimal performance and delivering coherent guidance to the sales representatives. Robust agent orchestration is therefore key to realizing the full potential of a multi-agent system.

7.3. Multi-Modal Data Processing Challenges

Handling a variety of data modalities, including text, images, and tables, requires the utilization of specialized AI models and sophisticated processing techniques tailored to each format. Furthermore, achieving a consistent semantic understanding across these diverse modalities can be a complex undertaking. Implementing effective multi-modal capabilities demands expertise in managing diverse data formats and ensuring a unified interpretation of their content.

7.4. Security and Compliance in Healthcare

In the healthcare sector, the protection of sensitive patient information and proprietary company data is of utmost importance 6. The multi-agent multi-modal RAG solution must be designed and implemented to comply fully with all relevant regulations, such as HIPAA in the United States, and other applicable data privacy laws 6. Security and compliance considerations must be integrated into every stage of the development and deployment process to ensure the confidentiality, integrity, and availability of sensitive data.

7.5. User Adoption and Training

The successful adoption and effective utilization of the new solution by the sales representatives are crucial for achieving the desired business outcomes. This requires providing comprehensive training and ongoing support to

ensure that sales teams are comfortable and proficient in using the system. Highlighting the tangible benefits of the solution and proactively addressing any concerns or potential resistance to change are important steps in fostering widespread user adoption.

8.Orchestrating Agents for Optimal Sales Guidance

To ensure the multi-agent system effectively delivers optimal sales guidance, a well-defined orchestration strategy is essential.

8.1. Defining Agent Roles and Responsibilities

The first step in effective orchestration is to clearly define the specific tasks and responsibilities of each agent within the multi-agent system. This includes a precise understanding of the functions of the Knowledge Retrieval Agent, the analytical capabilities of the Strategy Recommendation Agent, the content creation duties of the Content Generation Agent, and the supervisory role of the Orchestrator Agent. Clearly delineated roles prevent functional overlap and ensure that each agent contributes optimally to the overarching goal of providing comprehensive sales guidance.

8.2. Designing the Agent Workflow for HCP Interactions

Mapping out the step-by-step process through which the agents will interact to address a sales representative's query is crucial for efficient operation. This workflow design should take into account various interaction scenarios, such as the information needs during pre-call planning, the types of questions that might arise during an HCP meeting, and the data required for post-call follow-up activities. A well-designed workflow ensures that the agents collaborate seamlessly and efficiently to deliver timely and relevant guidance at each stage of the sales process.

8.3. Implementing Communication Protocols Between Agents

Establishing clear communication protocols that govern how the agents will exchange information and coordinate their actions is vital for the smooth functioning of the multi-agent system. This may involve implementing specific message formats, utilizing shared memory resources, or employing other inter-agent communication mechanisms provided by the chosen orchestration framework. Effective communication is fundamental for enabling the agents to collaborate efficiently, resolve conflicts, and avoid delays in providing guidance.

8.4. Incorporating Decision-Making Logic and Reasoning

Equipping the Strategy Recommendation Agent with sophisticated decision-making logic and potentially integrating machine learning models is essential for its ability to analyze retrieved information and provide insightful and actionable recommendations. This may involve the implementation of advanced reasoning techniques such as chain-of-thought prompting or the application of reinforcement learning methodologies to continuously improve the quality of strategic advice provided. The sophistication of this decision-making logic directly influences the intelligence and ultimate effectiveness of the strategic guidance offered to the sales team.

8.5. Monitoring and Evaluating Agent Performance

Implementing robust mechanisms to track the performance of each individual agent and the overall multi-agent system is a critical aspect of ensuring its ongoing effectiveness. This includes defining and monitoring key performance indicators (KPIs) such as response times, the accuracy and relevance of the information provided, and ultimately, user satisfaction among the sales representatives. Continuous monitoring and evaluation are necessary to identify areas for potential improvement, fine-tune the system's performance, and ensure that it continues to meet the evolving needs of the sales team.

9. Conclusion: Empowering Sales Teams with Intelligent Content Enablement

In conclusion, a multi-agent multi-modal RAG solution holds immense potential to revolutionize content enablement for sales representatives within the healthcare sector. By seamlessly integrating advanced retrieval and generation capabilities with the collaborative power of multiple specialized AI agents, this approach offers a transformative solution for enhancing HCP interactions. The ability of the system to provide both comprehensive knowledge, drawing from diverse digital content formats, and strategic guidance, tailored to the specific context of each interaction, equips sales teams to engage with HCPs more effectively and confidently.

The key benefits of implementing such a solution are manifold, including significant increases in sales productivity and effectiveness, improved engagement and satisfaction among HCPs, enhanced operational efficiency and reduced costs, strengthened compliance with industry regulations, and the inherent scalability and adaptability of the system to future needs. While the implementation of a multi-agent multi-modal RAG solution presents certain challenges, particularly in the areas of data integration, agent orchestration, multi-modal data processing, security and compliance, and user adoption, these can be effectively addressed through careful planning, expert execution, and a commitment to continuous improvement.

Looking ahead, AI-powered solutions like this multi-agent multi-modal RAG system will play an increasingly vital role in empowering sales teams within the healthcare industry. By providing intelligent content enablement, these systems will drive greater success in HCP interactions, foster stronger relationships, and ultimately contribute to improved patient outcomes in the evolving healthcare landscape.

References

Jain, N., Agrawal, S., Salunkhe, T. (2024). Identifying the Right Key Opinion Leaders in Pharma Using Machine Learning: Creating a Data-Driven Process for Impactful Outcomes. *Journal of the Pharmaceutical Management Science Association*, 5-16.

Gupta, A. (2024). Physician Engagement Optimization: Reinforcement Learning-based Omni-Channel GenAI approach for Maximizing Email Open Rates and embracing Representative preferences to target HCPs. *Journal of the Pharmaceutical Management Science Association*, 17-23.

Zhu, J., Sukumar, R., Majumder, A. (2024). Improve Customer Experience and Omnichannel Effectiveness through Customer Journey Analytics. *Journal of the Pharmaceutical Management Science Association*, 24-33.

Chehoud, C., Meng, X., Kumar, M., Chafekar, K., Singh, M. (2024). Optimizing Launch Excellence: An AI-driven Framework for Precision Engagement. *Journal of the Pharmaceutical Management Science Association*, 34-42.

Mitra, A., Sharma, A., Das, A., Sunder, A., Chilukuri, S., Trivedi, S. (2024). Transfer learning approach to enhanced patient classification using Real World Data. *Journal of the Pharmaceutical Management Science Association*, 43-58.

Parmar, R., Bawa, R., Khandelwal, H., Singh, V., Devi, S. (2024). Navigating Uncertainty: Evaluating Risks to Enhance Drug Sales Forecast. *Journal of the Pharmaceutical Management Science Association*, 59-73.

Ramesh, S., Karuppusamy, K. (2024). Healthcare Provider (HCP) Behavior Assessment: Identifying latent subgroups of HCPs and Salesforce eSales Aid Impact Analysis. *Journal of the Pharmaceutical Management Science Association*, 74-89.

Tsang, J. P., Larroque, M. (2024). Data X Ray - A Novel Tool that Assesses Data Sources Using ChatGPT. *Journal of the Pharmaceutical Management Science Association*, 90-102.

Gunasekaran, A. B., Nabha, N. (2024). Unleashing the Power of Rep Notes: Extracting Actionable Insights through NLP driven Analysis with LLMs and Generative AI. *Journal of the Pharmaceutical Management Science Association*, 103-112.

Sun, R., Kamin, S. (2024). Effectively predict patient discontinuation with AI and opportunities for Rx switching. *Journal of the Pharmaceutical Management Science Association*, 113-117.

About the Authors

Shihan He is the Machine Learning Engineer of the AI Center of Excellence at Novo Nordisk, where she specializes in the development and implementation of advanced AI solutions that drive decision-making throughout the healthcare value chain. With expertise in creating sophisticated AI systems, Shihan is dedicated to utilizing AI to enhance patient outcomes, particularly for those managing chronic diseases. Her work emphasizes the exploration and application of Generative AI (GenAI) to build scalable and robust AI applications that address critical Pharma challenges.

Anubhav Srivastava is the Associate Director of the AI Center of Excellence at Novo Nordisk, where he leads the vision, strategy, development, and deployment of AI-powered solutions across the organization. With extensive 19+ years' experience in building and managing high-performing teams of Data, Integration, Analytics, BigData, Robotics Process Automation and various technology platforms, Anubhav is dedicated to fostering a culture of innovation and collaboration within the organization. He is committed to driving the adoption of cutting-edge AI technologies, including GenAI, MLOps and LLOps, to tackle critical business challenges and enhance better patient outcomes.

Unlocking Commercial Success with Multimodal LLMs: Approach and Comprehensive Validation Framework

Daniel Young, AstraZeneca; Atharv Sharma, ZS Associates; Arindam Paul, ZS Associates; Prashant Mishra, ZS Associates; Arrvind Sunder, ZS Associates

Abstract: The increasing use of Large Language Models (LLMs) in healthcare commercial and medical applications, presents significant validation challenges owing to highly regulated nature of the industry. Validation is particularly difficult when the response is derived through a Chain-of-Thought (CoT) process querying healthcare databases, such as EMR claims data which itself carries inherent issues, such as inconsistent and incomplete capture of clinical events.

We encountered these challenges in the "LC-ChatIQ" agent developed to answer user's question, by querying on a backend database consisting of EMR claims and brand sales information. In this context, final response by LLM agent could be impacted by either the inherent data related issues at the source level or LLM's incapability in reasoning the correct Chain-of-Thought and resultant queries.

To mitigate such challenges, we propose a validation framework that separates data confidence from LLM action confidence. Data confidence is assessed based on source reliability, completeness, and consistency, while LLM action confidence focuses on the complexity of reasoning and query formulation. LLM confidence is derived using a guard-rail logic that breaks actions into attributes such as reasoning steps and query complexity.

Unlike existing methods that rely on real-time validation and improve LLM reasoning, our framework efficiently addresses both data and action confidence without requiring additional LLMs. The framework was tested on a validation response database of 100 user questions, with 90 valid outcomes. The results showed that the framework met success criteria in identifying drug/disease settings, information granularity, and recency, with an 82% success rate in assessing LLM action attributes. The underlying approach within our proposed framework allows for improved performance in agentic setting and scalability to other therapeutic areas.

1. Introduction

The integration of Large Language Models (LLMs) into healthcare is rapidly transforming the field, unlocking new possibilities across a wide range of applications. From enhancing research in oncology and rare diseases to supporting drug discovery, medical imaging, and commercial functions, LLMs are becoming indispensable tools. However, the highly regulated and precision-driven nature of the healthcare sector means that the application of LLMs must undergo strict validations to ensure their reliability and accuracy, since healthcare decisions based on LLMs, (or any insentient intelligence tool for that matter) can have profound consequences.

One of the primary challenges in LLM validation arises from the absence of 'source of truth' against which LLM responses can be measured. The issue is further worsened when

LLM responses are generated through Chain-of-Thought based reasoning, where the agents query healthcare-specific databases such as Electronic Medical Records (EMRs) data. In such cases, the integrity of the underlying data itself must also be considered, as healthcare databases may have issues related to under capture, capture biases, completeness, and granularity, making the validation a multifaceted challenge. This means that LLMs can produce inaccurate results if the database or reasoning process is flawed.

These challenges were encountered in our development work of LC-ChatIQ (Figure 1), an LLM agent designed to support cross-functional teams, HQ leadership and field representatives. The backend of LC-ChatIQ integrated a variety of data sources, each with its own intricacies and limitations such as varying granularities, data incompleteness, and variational capture, which could potentially affect the accuracy of the agent's final responses. Furthermore, the Chain-of-Thought process required by the agent introduced additional validation challenges.

Given these set of validation complexities, we required a more nuanced framework to assess the reliability of the LLM responses, one that could differentiate between errors originating from the data itself and those stemming from the agent's reasoning process.

To address this, we introduced two distinct categories of confidence: Data Confidence and LLM Action Confidence. Data Confidence relates to the quality and reliability of the data inputs - how accurate, complete, and consistent the underlying information is. LLM Action Confidence, on the other hand, focuses on the model's ability to process and reason through the data effectively.

2. LC-ChatIQ elements

LC-ChatIQ at its core is a suit of agentic LLM module implemented using the ReAct[1] (Reasoning and Acting, proposed by Zhang et al., 2023) framework. Below are core key components of LC-ChatIQ:

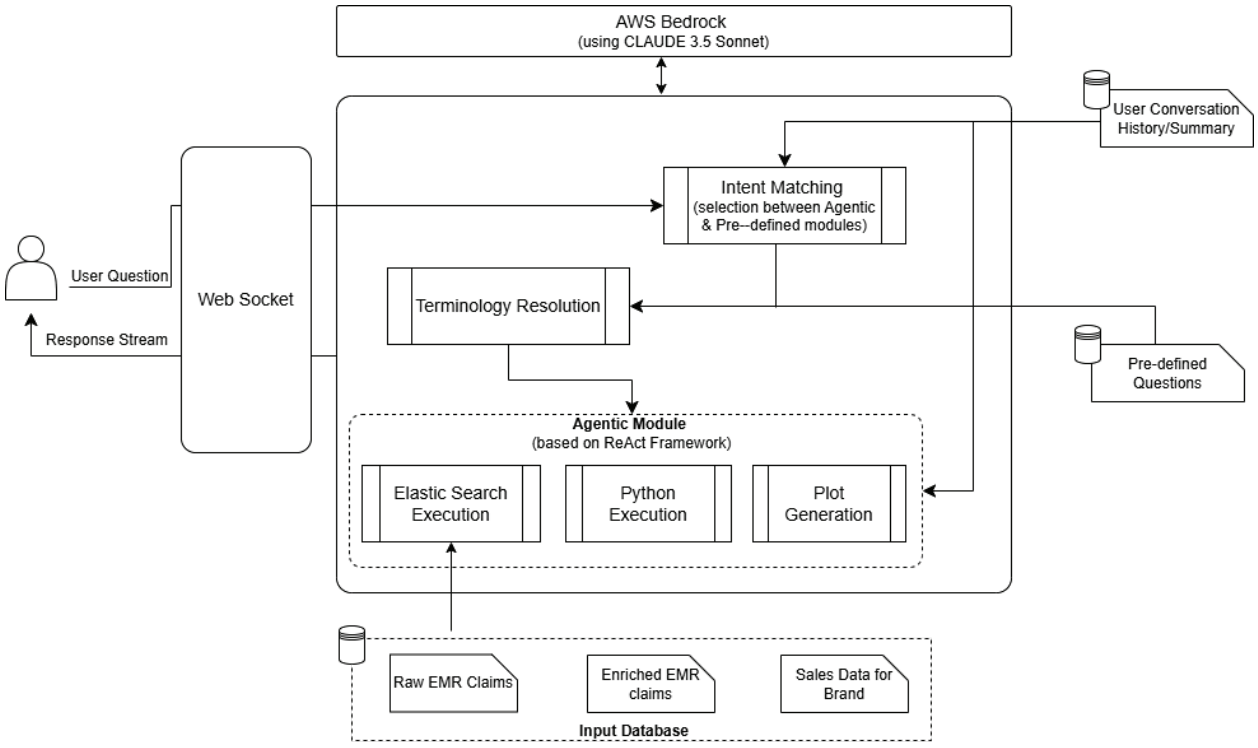


Figure 1: LC-ChatIQ Agent Architecture View

Database: In the presented version of LC-ChatIQ datalake, there are 3 major sources of information for agent to derive their responses from. The data information and their consolidation into final database is presented below:

LC-ChatIQ Database Consolidation				
	Source Type	*Description*	*Sourced Information*	*Data Granularity*
Data Sources	Synthetic EMR claims[1]	Contains machine learning based Synthetic Dx, Px, Rx information on top of existing information available in Raw EMR claims.	Diagnoses/Procedures	Patient
			Prescription pattern	
			Drug drop off	
			Duration of Therapy	
			Biomarker information	
	Raw EMR claims	Contains Dx, Rx, Px information at patient level, although under-captured compared to known benchmarks.	Treatment outcomes	Patient
			HCP interactions	
			Referral distribution	
	Sales Distribution		NPS / TRx data of Orals	HCP
		Sales information of Brand drugs	NPS / TRx data of IVs	Account
Granularity affiliation	Patient —*Rule based*→ HCP —*Rule based*→ Parent —*Direct*→ Top Parent			
Patient statistic aggregation	*Granularity level*: HCP, Parent, Top Parent			
	Interval level: Weekly / Monthly			

[1] Synthetic EMR claims is Raw EMR claims enhanced with AI generated Dx / Px / Rx claims which provides more comprehensive representation of the real-world disease landscape.

Table 1: Raw database consolidation

The primary data sources under consideration include structured tables at various granularities. These sources primarily encompass EMR claims (both Raw and Synthetic), with data available at the patient level, and Sales Data for branded drugs, with data granularity at the HCP level for oral drugs or account level for IV drugs.

The final consolidated database contained weekly/ monthly level patient statistics aggregated at HCP or Account (Parent and TopParent) level, which are sourced from EMR sources and Sales Data. Since the source level information is present at different granularities, the appropriate mapping mechanism was applied to level up the information at other higher granularities (shown in granularity affiliation section in table 1). The patient statistics are relevant for numerous key indications/ trials in Lung Cancer as well as different patient journey points like Diagnosis, Surgery, Biomarker tests etc.

Terminology Resolver: The Terminology Resolver is a crucial component in the LC-ChatIQ architecture, designed to enhance the LLM agent's understanding and response accuracy, especially in domain-specific contexts. for e.g Biomarker Testing rate, NPS are domain specific terms that LLM agent might not have full understanding of. Resolver interprets and clarify domain-specific terms and defintions, jargon, acronyms, and specialized vocabulary that may appear in user queries.

Agentic Module (ReAct framework): Once the Synthetic user query from the Terminology Resolver is processed, The ReAct based method decomposes the problem into distinct, logical steps, each involving code-driven operations to handle tabular data as needed.[2] The context from User Conversation History/Summary aids to module's functionality for personalized and contextually relevant responses.

3. Related Work

Herein, we primarily focus on the published LLM applications in healthcare domain and delineate the approach adopted for validating/ evaluating the agentic response and their crucial impact if any, on the use case objectives. We further examine whether the considered evaluation criteria within the work, had any source of truth or existing context to aid the validation methodology.

Du et al.[3] conducted a comprehensive review of 76 published articles to assess the application of LLMs in analyzing real EHR data for enhancing patient care. Their findings indicate that the integration of multimodal EHR data with LLMs remains rare. Although, they concluded that agentic workflows over structured EHR data can significantly improve decision-making and facilitate more accurate diagnoses, particularly for rare diseases, but also underlined that the major limitation which hinders the adoption of LLMs in healthcare setting, is the lack of standardized evaluation methods. Within the reviewed studies, the most common evaluation metrics were correctness (56 studies), agreement with experts or ground truth (12 studies), and completeness, reliability, stability, and readability (each in 7 studies). For accuracy assessment, confusion matrix-based metrics were predominantly used.

Gilbert et al.[4] attempted to automate the extraction of critical health information for cancer patients by searching for keywords in a patient's EMR using an LLM. Prompt engineering was used to iteratively optimize the output, which was then compared to ground truth data that was abstracted by manual review. The effectiveness of the LLM was quantitatively assessed by calculating precision, recall, accuracy, and F1 score.

Nachane et al.[5] employed human evaluation techniques to categorize the correctness of agent responses derived on a clinical MCQ database, utilizing a 5-shot Chain-of-Thought prompting strategy. A reward-based mechanism was integrated to fine-tune the agentic performance.

Cui et al.[6] introduced the EHR-CoAgent system, a two-agent LLM framework designed to predict diseases from EHR data while investigating the feasibility of applying Large Language Models (LLMs) to convert structured patient visit data (e.g., diagnoses, labs, prescriptions) into natural language descriptions. This system includes a predictor agent that makes predictions and explains the reasoning, and a critic agent that reviews incorrect predictions and offers corrective feedback. The critic agent's feedback is used to update the prompts given to the predictor agent, enabling the system to learn from its mistakes and adapt to the specific challenges of the EHR-based disease prediction task.

Sudarshan et al.[7] explored methods for generating patient-friendly radiology reports via an LLM-driven multi-agent workflow. This approach aims to reduce the necessity for medical professional verification. They evaluated agent responses using readability and accuracy metrics, employing the Flesch-Kincaid grade level (aimed at a US 6th-grade reading level) and ICD-10 code matches, respectively.

The consolidated view of the literature research is provided below with information and learnings relevant to the current use case.

Authors	Healthcare use-case	Used data	LLM elements	Presence	Ground truth	Validation criteria
Cui et al.	Disease prediction	EHR (Dx, Px, Rx)	CoT/Reasoning	Yes	_	_
			Response	Yes	Yes	F1 score
Du et al.	Miscellaneous, Systematic Review (n = 76)	EHR	CoT/Reasoning	In some (n=3)	_	_
			Response	Yes	Yes	Correctness, Manual confirmation
Nachane et al.	Medical question answering	MEDQA-MCQ	CoT/Reasoning	Yes	_	_
			Response	Yes	Yes	Manual confirmation
Gilbert et al.	Care information extraction	EMR clinical notes	CoT/Reasoning	No	_	_
			Response	Yes	Yes	Manual confirmation
Sudarshan et al.	Medical report summarization	Radiology report	CoT/Reasoning	No	_	_
			Response	Yes	Yes	Readability, Accuracy
Qu et al.	Medical event information extraction	EMR clinical notes & open-source MIMIC	CoT/Reasoning	No	_	_
			Response	Yes	Yes	Accuracy, Sensitivity, Specificity

Table 2: Literatures reviewed

To summarize the literature review on LLM applications in healthcare, key finding was that the majority of the literatures relied on zero-shot or few-shot prompting techniques, rather than incorporating Chain-of-Thought prompting. While a limited number of studies do employ Chain-of-Thought prompting, they frequently overlook any validation criteria for the reasoning steps involved, instead relying on the LLM itself to assess the nature of thought pattern. In certain instances, the effectiveness of Chain-of-Thought prompting is evaluated solely based on improvements in final performance, which inherently requires the presence of ground truth data by definition. On the other hand, for use cases like question-answering, information extraction or summarization, the response evaluation method was mostly manual confirmation and judgement.

4. Key considerations in LC-ChatIQ validation

One of the key issues identified in the reviewed literature is the absence of validation criteria for the underlying data, particularly given the inherent flaws in EMR, and clinical notes data. EMR or claims data often fail to cover a sufficiently diverse patient and healthcare provider population, limiting the ability to draw confident conclusions. Additionally, these datasets are frequently compromised by the underreporting of critical medical events such as diagnoses, procedures, and treatments. Making critical decisions based on incomplete data, whether in commercial or medical applications, poses significant risks, especially within the highly regulated healthcare industry.

Another notable challenge highlighted during the exploratory work is the lack of evaluation criteria for Chain-of-Thought reasoning steps, particularly in unsupervised settings where ground truth data may be limited or unavailable. Relying on another LLM to assess the reasoning steps does not fully address the issue, as it introduces the need to manage yet another model. This necessitates ensuring that the secondary LLM is provided with appropriate instructions and data, while also adding additional complexity, cost, and resource demands to the system.

Therefore, the key challenges realized in LC-ChatIQ validation are as follows:

- How do we validate LLM response in **Absence of Source of Truth** for User questions?
- How do we validate **underlying Chain of Thought and Data Sources** used?
- How do we ensure **explainability** of LC-ChatIQ confidence comprehensible to the end user?
- Can the validation framework be made **scalable** as LC-ChatIQ application expands across Data sources and Tumors?

5. Validation Criteria

To address the challenges posed on LLM validation within our LC-ChatIQ application, we propose a robust, two-pronged validation framework that evaluates both **data confidence** and **LLM action confidence**.

5.1 Data Confidence

The first component, **data confidence**, assesses several key aspects related to the underlying database: the reliability of data sources, the completeness of the data, consistency in the temporal capture of events, and the patient-healthcare interaction mapping. Data confidence types along with their corresponding key considerations are mentioned below:

Confidence type	Considerations
Data Source Confidence	• Data sources considered the "source of truth" (e.g., Sales Data, Raw Claims) generally exhibit a high degree of confidence in their accuracy. • Data sources directly referenced by business units or headquarters typically carry a high level of confidence. • The confidence level in sources containing synthetic data (e.g., Synthetic Claims) varies, depending on the method of data synthesis, ranging from low to high.
Data Completeness / Coverage	• Sources with significant undercoverage of patients, healthcare providers, or accounts (e.g., Sales data for non-TRx claims, Claims data with incomplete diagnoses or treatments) tend to have low confidence.
Temporal Capture	• Recent analyses of claims data generally exhibit lower confidence due to the inherent limitations of recent data capture.
Entity Affiliation	• Data sources with rule-based entity mapping (e.g., HCP, parent, top-parent affiliations) often result in diluted confidence, depending on the success rate of the mapping process (e.g., Claims data linking patients to accounts for insights). • Child-Parent-Subnational mappings typically have high confidence due to their direct nature, while Patient-HCP-Child mappings are subject to diluted confidence because they rely on rule-based methods.

Table 3: Data confidence elements

5.2 LLM Confidence

On the other hand, **LLM action confidence** focuses on the agent's reasoning processes and its approach to query formulation, which is quantified through analysis CoT complexity and resultant query meta data. These are presented in the table below:

Confidence Type	Criteria	Reasoning
Reasoning Complexity	Number of reasoning steps	• A greater number of reasoning steps is expected to correlate with a lower level of confidence, as the system must process more complex logic to address intricate queries.
Query Complexity	Number of Aggregations Number of Filters	• The complexity of a query, as indicated by the number of aggregations, filtrations, or subquery nesting, is likely to reduce confidence in the result.
Columns referred	Number of Indices referred Number of Columns referred	• The greater the number of indices referenced, the higher the probability of failure in producing an accurate answer. • A larger number of columns referenced in a query typically correlates with a lower degree of confidence in the results.
Join steps across tables	Number of intermediate tables created Number of Merges/Joins applied	• An increased number of joins between intermediate tables tends to diminish the likelihood of arriving at a correct final answer.

Table 4: LLM confidence elements

The complexity of LLM action is determined through a historical response database containing over 1,600 use-case specific questions and corresponding agent responses. These responses span a spectrum from direct to open-ended queries and are used to define a median complexity for actions. By measuring deviation from this median, we can effectively score the LLM's reasoning performance.

5.3 Validation Quantifiability

	Confidence criteria	Quantifiable?	Scoring criteria
Data Confidence	Source Confidence	Yes	Confidence within Data collection (raw / synthesized / processed)
	Data Completeness	Yes	Longitudinal & Lateral coverage rate of data points compared to expected/benchmark
	Temporal Consistency	No	Temporal data capture varies with the type of captured clinical event
	Affiliation Confidence	No	Qualitative confidence within affiliation mapping
LLM Confidence	Reasoning complexity	Yes	Deviation of aggregated criteria against Median/expected
	Query complexity	Yes	
	Columns referred	Yes	
	Join steps	Yes	

Table 5: Validation quantifiability assessment

6. Proposed Validation Framework

The following outlines a proposed general-purpose framework for CoT-based Agentic Response. In this framework, the response can either be the direct answer or a runnable query/function based on the underlying data. Through our literature review, we found that when the response is a direct answer, "human/expert confirmation" is necessary for accurate response evaluation. In such cases, either a ground truth or a reasonable approximation of the ground truth is required to assess the agent's response.

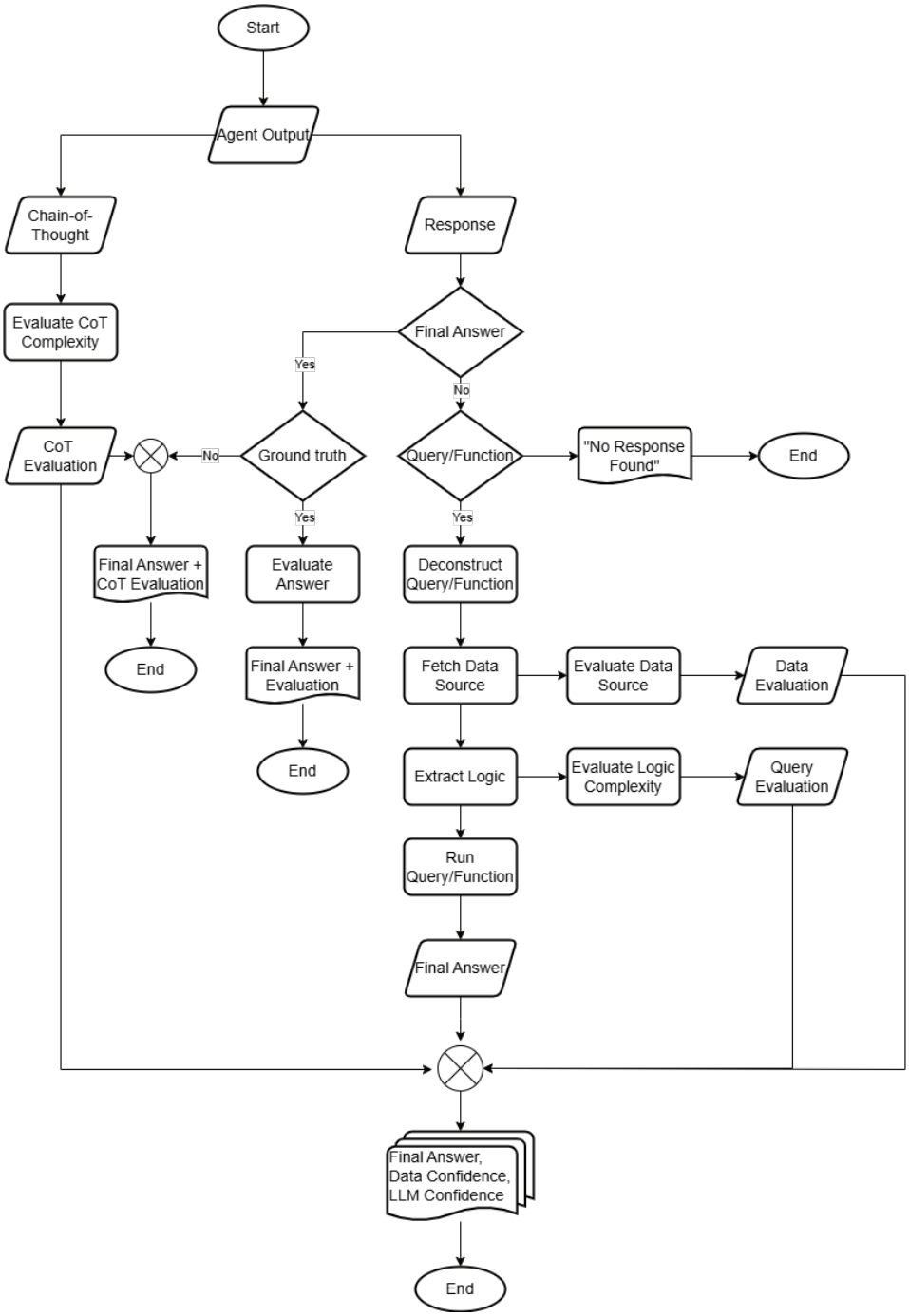


Figure 2: Proposed general purpose validation framework

In the absence of a clear ground truth, CoT evaluation can be derived based on specific assessment criteria. Adlakha et al.[8] demonstrated that instruction-following question-answering agents tend to underperform when handling open-ended or elaborated questions, with elaborated answers being the most common failure category. In CoT-based systems, (1) for complex or elaborated questions, the resulting CoT tends to be longer and more intricate than average, which can lead to factually incorrect final answers; (2) for open-ended questions, the agent may make assumptions, where a false assumption could result in an incorrect answer. We further test this hypothesis before developing the CoT evaluation strategy.

6.1 "Complex questions lead to complex Chain-of-Thought & querying steps"

To test this hypothesis, we examined a diverse set of questions drawn from historical question records. Human judgment was employed to assess the complexity and nature of each question, categorizing them into three distinct buckets. Additionally, we analyzed the associated Chain-of-Thought (CoT)/Queries and final responses to better understand the relationship between the question, reasoning steps, and the resulting answer. These three buckets are tailored to the specific types of questions LC-ChatIQ is likely to encounter, as well as the level of reasoning and calculation the LLM Agent needs to perform, based on the provided context. The categorization is based on the effort required by a human to answer similar questions, given the same context.

- **Direct Questions:** These questions typically involve a limited number of steps and require relatively straightforward reasoning.
- **Calculation-Intensive Questions:** These questions demand a higher degree of calculation, though the steps involved can be predefined, regardless of their complexity.
- **Open-Ended/Inferential/Out-of-Scope Questions:** These questions are more abstract, requiring inference based on observation. In some cases, the entire chain of reasoning may be predetermined, while other questions may fall outside the agent's scope or expertise.

Question type	Examples
Direct	– top 5 accounts for taggriso? – what is the top account for adaura patient volume – top 5 accounts with it's account id ? – can you show labels by indication for all AZ drugs – what is the top subscriber for Tagrisso – …
Calculation Intensive	– which accounts imfinzi marketshare have declined in last 6 months ? – What is the monthly trend of Tagrisso share for FLAURA? – which accounts perform best and which are the worst – what is flaura market share over time – …
Open-Ended/Inferential/OoS	– is there a correlation between adaura biomarker testing rate and tagrisso share? – High volume HCPs who are shifting to IO treatments for FLAURA and their biomarker testing rates? – …

Table 6: Question examples corresponding to their type

To further explore the questions, they were processed through LC-ChatIQ to extract the corresponding Chain-of-Thoughts (CoT)/Queries and Final Responses. A detailed analysis was conducted on key metadata elements, such as the number of Queries, Columns, Aggregations, and Filters associated with these questions. Additionally, we performed a pairwise Tukey HSD (Honest Significant Difference) test[9] across the distinct question buckets to assess the statistical significance of the differences observed

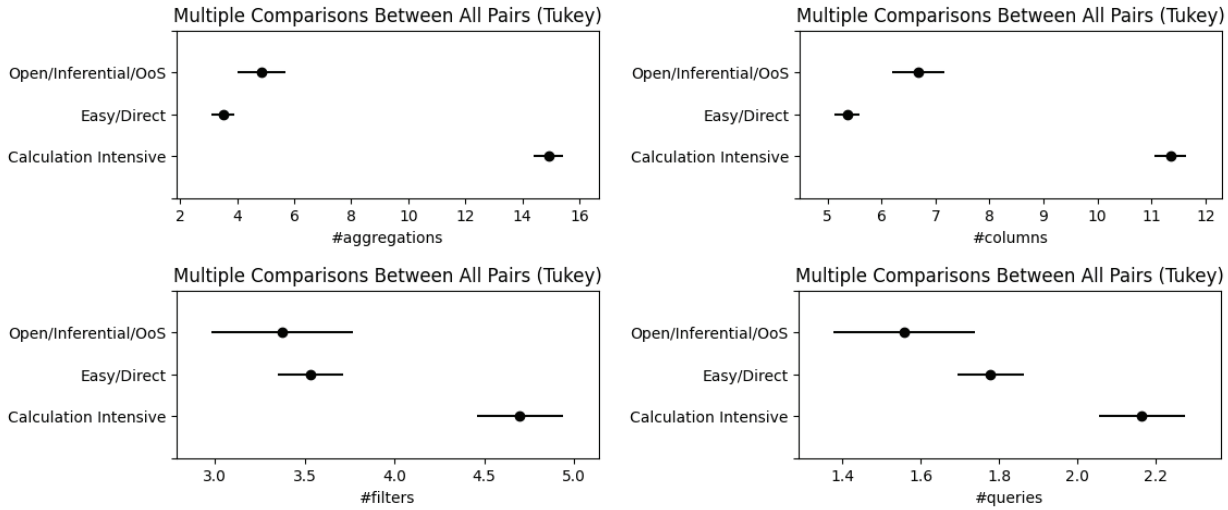


Figure 3: HSD plots for CoT elements across question groups

Metadata	Group1	Group2	Mean-diff	p-adj	Lower	Upper	Reject
#Aggregation	Calculation Intensive	Direct	-11.4127	0.0	-12.3194	-10.5061	True
	Calculation Intensive	Open/Inferential/OoS	-10.0738	0.0	-11.4279	-8.7198	True
	Direct	Open/Inferential/OoS	1.3389	0.0289	0.1093	2.5685	True
#Columns	Calculation Intensive	Direct	-5.9904	0.0	-6.5114	-5.4693	True
	Calculation Intensive	Open/Inferential/OoS	-4.6749	0.0	-5.4531	-3.8967	True
	Direct	Open/Inferential/OoS	1.3155	0.0	0.6088	2.0221	True
#Filters	Calculation Intensive	Direct	-1.166	0.0	-1.5904	-0.7416	True
	Calculation Intensive	Open/Inferential/OoS	-1.3224	0.0	-1.9562	-0.6886	True
	Direct	Open/Inferential/OoS	-0.1565	0.7991	-0.732	0.4191	False
#Queries	Calculation Intensive	Direct	-0.3864	0.0	-0.581	-0.1918	True
	Calculation Intensive	Open/Inferential/OoS	-0.6062	0.0	-0.8969	-0.3156	True
	Direct	Open/Inferential/OoS	-0.2198	0.1241	-0.4837	0.0441	False

Table 7: HSD performance table across CoT meta data

Note: Since we are conducting Pairwise comparisons, Group1 and Group2 are just the paired groups of 'Question types' for a certain metadata (eg: #Aggregation) comparison. For example, first row in the table above represents the statistical comparison of #Aggregation across 'Calculation Intensive' and 'Direct' question types.

Based on the HSD test and plots, we can draw few key inferences:

- **Calculation-intensive problems** generally require a slightly higher number of reasoning steps on average, but they involve significantly more aggregation, filtration, and column selection steps to derive the answer, compared to **Direct or Open/Inferential/OoS** questions.

- **Direct** and **Open/Inferential/OoS** questions primarily differ in terms of aggregation and column selection. **Open/Inferential/OoS** questions present slightly more complexity for the LC-ChatIQ, but further analysis reveals that the LLM agent typically handles OoS questions with relative ease, providing a fitting response without much reliance on calculation. In contrast, for Open/Inferential questions, the LLM agent often resorts to calculation steps that may not directly address the user's query. However, when the answer is reasonable, the calculation steps tend to be straightforward, much like those seen in **Direct** questions.

6.2 "Chain-of-Though complexity leads to diluted accuracy in final answer"

To test this premise, we evaluated LC-ChatIQ's responses across multiple iterations of 62 test questions (refer appendix). With the guidance of a business expert, we manually analyzed how often LC-ChatIQ was able to generate reasonably correct responses. To assess the consistency of its accuracy, we randomly asked the same set of questions in four different sessions, gathering LC-ChatIQ's responses for evaluation. Below are the questions used for assessing LC-ChatIQ's performance:

Difficulty is defined based on the following criteria:

- **Easy:** LC-ChatIQ correctly identifies the answer in at least 3 out of 4 sessions.

- **Medium:** LC-ChatIQ identifies the answer at least once, but no more than twice, out of 4 sessions.

- **Tough:** LC-ChatIQ fails to identify the correct answer in any of the 4 sessions.

Based on the final evaluation, the assessment grid is provided below:

	Difficulty →		
Cluster ↓	Easy	Medium	Tough
Direct	11	2	1
Calculation Intensive	18	10	0
Open/Inferential/OoS	15	3	3

Table 8: Question cluster vs difficulty faced by LC-ChatIQ

From the grid, it is evident that LC-ChatIQ struggles most with Open/Inferential/OoS type questions. In contrast, for Calculation Intensive questions, the agent is more likely to provide correct answers, although not in every instance. While the agent may face difficulty in arriving at the correct answer on every attempt with calculation-intensive queries, it still demonstrates a stronger ability to reason correctly, select the appropriate schema, and generate the right queries to arrive at the correct conclusion

Now that we establish the above two premises i.e. (1) Complex questions result in complex chain-of-thoughts, (2) Complex chain-of-thought results in higher likelihood of inaccurate response, we use the Chain-of-thought complexity elements to pseudo evaluate the LLM confidence. (Note: Since we would not

have availability of ground source of truth in real time for direct evaluation of LLM response, assessment of intermediate CoT complexity acts just as a proxy for LLM evaluation, hence this evaluation mechanism is termed as 'Pseudo Evaluation'). Chain-of-thought complexity is determined by combination of metadata information extracted from the underlying reasoning and querying steps, with elements which are significantly different across different question types weighted higher (e.g. #Aggregations, #Columns) compared to the ones (#Filters, #Queries) which don't exhibit significant differences across questions groups.

7. LC-ChatIQ Scoring Formulation

Given that the scope is focused on oncology, i.e. lung cancer, we developed a rule-based mechanism to map LC-ChatIQ's chain-of-thought reasoning to the associated data, schema, and sources. Additionally, the resulting queries can be deconstructed to extract key metadata, such as aggregations, columns, filters, and timeframes. Figure 4 represents the transformation flow of ChatIQ response into Data and LLM confidence.

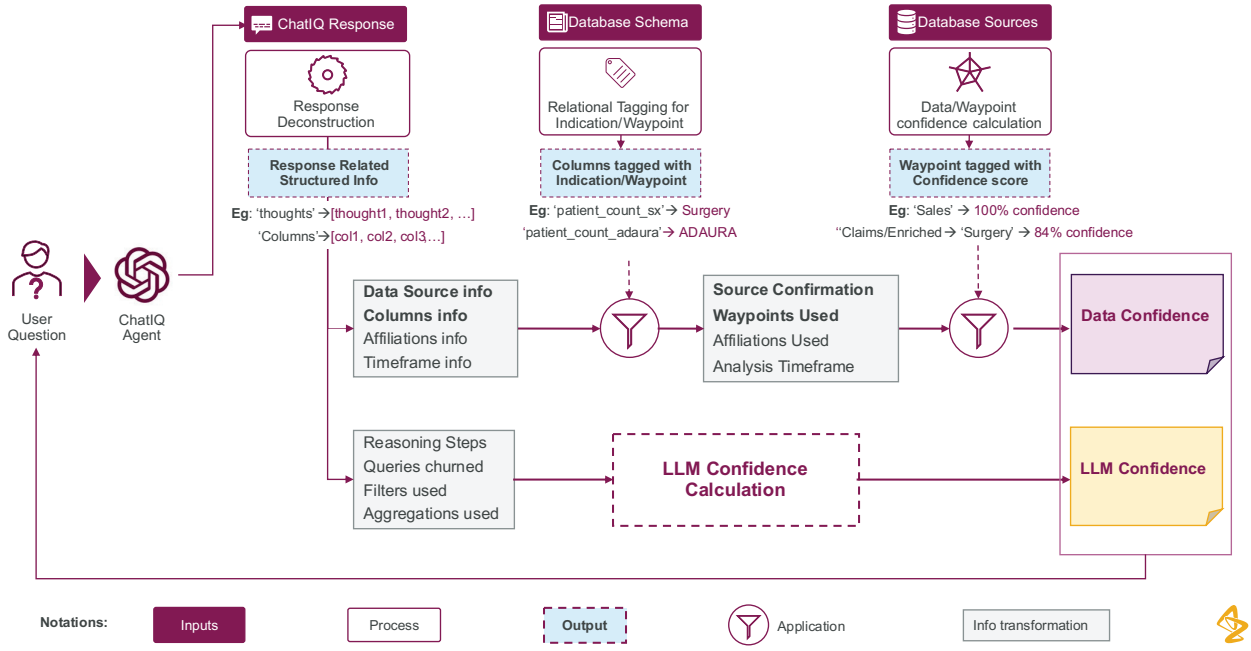


Figure 4: LC-ChatIQ response transformation into Data and LLM confidence

Table below shows an example of how the Chain-of-Thought is scraped to extract elements which help derive the Data and LLM confidence.

Question:	**Biomarker testing rate in NY?**	
CoT steps	***LLM response elements***	***Deconstruction***
Reasoning Steps	Certainly! I'll break down the process to calculate the biomarker testing rate for New York (NY) using the available data. Let's approach this step-by-step. To calculate the biomarker testing rate in NY, we need to: 1. Filter the data for NY state and aggregate at the Parent level. 2. Sum the total number of Flaura-eligible patients and the number of patients who received EGFR testing. 3. Calculate the biomarker testing rate as a percentage. Let's start with querying the necessary data: ...	# CoT Steps

Queries	Action Input: ({
	"size": 0,
	"query": {
	"bool": {
		# Filters
		# Aggregations
		# Columns
		# Queries

	"filter": [
	{"term": {"state": "NY"}},
	{"term": {"type": "Parent"}},
	{"term": {"timeinterval": "Monthly"}},
	{"range": {"intervalstart_date": {"gte": "now-18M/M", "lt": "now-6M/M"}}}
	]
	}
	},
	"aggs": {
	"total_patients": {
	"sum": {"field": "patient_count_flaura"}
	},
	"tested_patients": {
	"sum": {"field": "egfr_test_flag_count_flaura"}
	}
	}
	}, "database")
	
	Below is the generated Elastic search result in csv format:
	====
	category,value
	total_patients,3759.0
	tested_patients,2412.0
	
	====
	
	biomarker_testing_rate = (tested_patients / total_patients) * 100 if total_patients > 0 else 0
	result = f"The biomarker testing rate in NY is {biomarker_testing_rate:.2f}%"
		Time range
		Source Info
		Affiliation level

Answer	The biomarker testing rate in NY is 64.17%	

Table 9: An example of ChatIQ response deconstruction to extract Data Confidence & LLM Confidence elements. Orange text imply query metadata contributing to LLM confidence, Violet highlighted texts help mine Data confidence elements.

When the question is asked to the LC-ChatIQ agent, the CoT based response and resultant query is deconstructed to extract, reasoning steps and other relevant query meta data such as aggregations, filters and columns (fields) used. Further we also scrape information such as time range and affiliation information to evaluate temporal consistency and affiliation confidence, which remain qualitative in nature. The extracted columns also help in the identification of the actual data source, and as a result pin down the data completeness and inherent capture confidence.

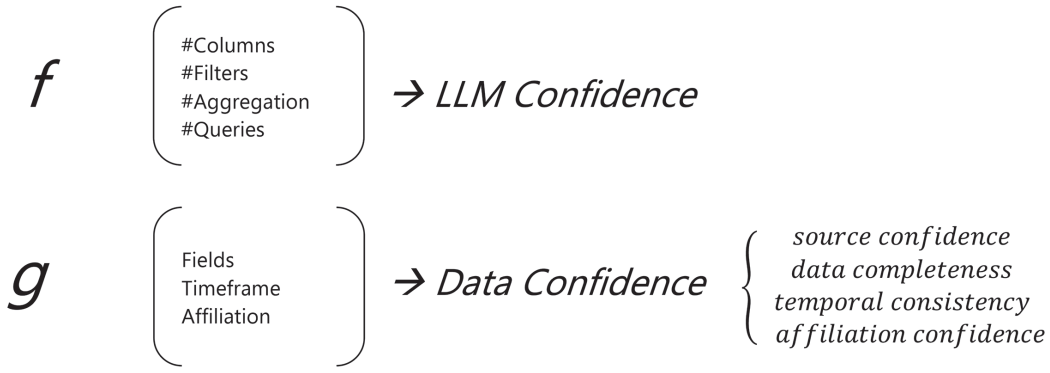


Figure 5: LC-ChatIQ evaluation scoring

8. Results

The effectiveness of the proposed validation approach was assessed through a manual evaluation of a validation response database comprising 100 user queries. The framework successfully produced the desired outcome for 90 of these queries, achieving a high success rate. However, for the remaining 10 responses, the agent encountered a recurring issue—entering into an endless correction loop that ultimately led to termination. The success criteria for the framework were defined as follows: (i) over 95% accuracy in the correct identification of the drug/disease setting, which is crucial for defining the market; (ii) over 90% accuracy in identifying the appropriate level of information granularity (e.g., National, State, Account, HCP); (iii) over 80% accuracy in identifying the recency of information, an important factor for data stability; and (iv) over 80% accuracy in identifying the relevant LLM action attributes, including elements like the number of Chains-of-Thought (COT), queries, columns, joins, aggregations, and filters.

Among the 90 valid outcomes, both **Data Confidence** and **LLM Confidence** scores were carefully analyzed for any discrepancies from both the data and generation perspectives. The first three success criteria were met with precision, as the guardrail methodology successfully fulfilled these requirements in all 90 cases. For the fourth criterion, the framework demonstrated a strong performance, achieving the target in at least 74 out of 90 cases (~82%) across all LLM action attributes.

# Questions	# Eligible Response	Confidence type	Success Criteria	# Accurate Identification	Score	Passed
100	90	Data Confidence	Identification of *Drug/Event*	87	97%	True
			Identification of *Granularity*	90	100%	True
			Identification of *Timeframe*	86	96%	True
		LLM Confidence	Identification of # *columns*	83	92%	True
			Identification of # *aggregation*	74	82%	True
			Identification of # *filters*	79	88%	True
			Identification of # *queries*	90	100%	True
	10 responses were ineligible					

Table 10: Evaluation performance table

9. Impact

From a usability standpoint, the confidence scores were accompanied by a 'confidence reasoning' feature that provided end users with a clear explanation of the rationale behind each score. This additional layer of transparency helped the leadership and cross functional teams with understanding clear sources of inaccuracies, particularly when complex data and model-generated responses are involved. Given that the scoring criteria are rules-based, we hypothesize that the framework would perform even more efficiently in real-time user-agent interactions compared to alternative LLM-based validation methods. However, a quantifiable performance boost in a real-world deployment setting has yet to be definitively measured.

One of the key strengths of this approach lies in its scalability. The generalizability of the LLM action confidence scoring system makes it adaptable for use across various domains and disease areas. This flexibility holds immense potential for scaling the proposed framework beyond the initial use case. Some of potential use cases of the presented evaluation mechanism across Medical and Commercial can be as follows:

9.1 Medical Use Cases:

- **Clinical Decision Support Systems:** In an LLM agent trained to assist healthcare professionals in making decisions based on medical data from EMR, the CoT based response ensures that the chatbot provides a logical, structured, and evidence-backed path that leads to a final recommendation. The evaluation can assess if the chain of thought is medically sound, logically consistent, and adheres to best practices.

- **Personalized Patient Care:** For personalized care, if the LLM agent is trained to guide healthcare providers in recommending tailored treatments and interventions, the CoT evaluation can ensure that the chatbot properly interprets patient data (e.g., age, medical history, medication adherence) and suggests personalized care pathways. It can also help prevent errors in treatment advice.

- **Clinical Trial Recruitment:** If an LLM is tasked to identify suitable candidates for clinical trials by analyzing a patient's health data and matching it with inclusion/exclusion criteria, the CoT evaluation ensures that the chain of thought leads

to the correct match without overlooking critical data points or generating inappropriate trial recommendations.

- **Patient Monitoring:** If an LLM is trained to raise attention alert for patients with chronic conditions or post-surgical care, based on their ongoing symptoms and treatment adherence, the CoT evaluation helps ensure that the reasoning behind an alert is sound and based on correct data from the EMR.

9.2 Commercial Use Cases:

- **Patient Engagement and Education:** If an LLM chatbot is trained to engage patients by answering questions about treatments, medications, and medical conditions, the CoT evaluation can ensure the chatbot's reasoning is accurate and the information provided is evidence-based.

- **Patient Opportunities and Targeting:** If an LLM agent is trained to extract diagnosis and treatment patterns from patient journeys to recommend subnational hotspots for targeting, the CoT evaluation can help quantify the accuracy of the reasoning behind the recommended hotspot and opportunity type.

- **Insurance Claim Processing and Fraud Detection:** If an LLM is tasked to assist in processing insurance claims, verifying patient eligibility, or identifying potential fraud by analyzing the medical records, the CoT evaluation ensures that the LLM's chain of thought aligns with regulations, best practices, and data integrity protocols to detect inconsistencies or errors.

10. Conclusions

Unlike existing open-source methods, which often rely on additional LLMs for real-time validation and focus on improving reasoning capabilities through multiple chains of thought and stepwise validation, our approach offers a more streamlined and efficient solution. These conventional methods often overlook the critical aspect of data confidence and tend to be computationally intensive, especially in production environments. Furthermore, they do not explicitly assess the complexity of LLM actions, which is essential in use cases like ours, where the multi-step nature of the agent's reasoning process makes it particularly susceptible to errors or hallucinations.

Our guardrailed framework overcomes these limitations by clearly distinguishing between **data confidence** and **LLM action confidence**, tackling both the evaluation of data integrity and the complexity of LLM reasoning. This differentiation is particularly crucial in healthcare applications, where the impact of inconsistent or incomplete data on decision-making can be profound. By isolating LLM action complexity, we ensure that the model's reasoning is evaluated independently of the underlying data, allowing for a more accurate and reliable assessment of the agent's overall performance.

11. Future Scope

To expand the framework to novel data sources, we anticipate the need of further considerations to score the data confidence. Hence, application of proposed framework would require new data sources to be carefully assessed and their confidence criteria to be pre-determined depending on the specific use case requirements. Furthermore, inclusion of other nuanced query/CoT metadata like number of error courses, similarity scores with assistant questions, can further inform the LLM confidence scoring mechanism, which can lead to more robust Chain-of-Thought evaluation.

References:

1. Zhang Y, Henkel J, Floratou A, Cahoon J, Deep S, Patel JM. ReAcTable: Enhancing ReAct for Table Question Answering. arXiv preprint arXiv:2310.00815. 2023 Oct 1.
2. Nahid MM, Rafiei D. TabSQLify: Enhancing Reasoning Capabilities of LLMs Through Table Decomposition. arXiv preprint arXiv:2404.10150. 2024 Apr 15.
3. Du X, Zhou Z, Wang Y, Chuang YW, Yang R, Zhang W, Wang X, Zhang R, Hong P, Bates DW, Zhou L. Generative Large Language Models in Electronic Health Records for Patient Care Since 2023: A Systematic Review. medRxiv. 2024 Aug 19.
4. Gilbert M, Crutchfield A, Luo B, Thind K, Ghanem AI, Siddiqui F. Using a Large Language Model (LLM) for Automated Extraction of Discrete Elements from Clinical Notes for Creation of Cancer Databases. International Journal of Radiation Oncology, Biology, Physics. 2024 Oct 1;120(2):e625.
5. Nachane SS, Gramopadhye O, Chanda P, Ramakrishnan G, Jadhav KS, Nandwani Y, Raghu D, Joshi S. Few shot chain-of-thought driven reasoning to prompt LLMs for open ended medical question answering. arXiv preprint arXiv:2403.04890. 2024 Mar 7.
6. Cui H, Shen Z, Zhang J, Shao H, Qin L, Ho JC, Yang C. LLMs-based Few-Shot Disease Predictions using EHR: A Novel Approach Combining Predictive Agent Reasoning and Critical Agent Instruction. arXiv preprint arXiv:2403.15464. 2024 Mar 19.
7. Sudarshan M, Shih S, Yee E, Yang A, Zou J, Chen C, Zhou Q, Chen L, Singhal C, Shih G. Agentic LLM Workflows for Generating Patient-Friendly Medical Reports. arXiv preprint arXiv:2408.01112. 2024 Aug 2.
8. Adlakha V, BehnamGhader P, Lu XH, Meade N, Reddy S. Evaluating correctness and faithfulness of instruction-following models for question answering. arXiv preprint arXiv:2307.16877. 2023 Jul 31.
9. Abdi H, Williams LJ. Tukey's honestly significant difference (HSD) test. Encyclopedia of research design. 2010 Mar;3(1):1-5.

About the Authors

Daniel Young is a Senior Director - Data Science & AI, AstraZeneca, based out of San Francisco office. With over 20 years of experience across healthcare and biosciences domain, Daniel is currently leading the development and application of advanced data science and GenAI capabilities within AstraZeneca's commercial and R&D functions.

Atharv Sharma is an Associate Principal with ZS based out of Philadelphia and is a core member of ZS's RWD and insights venture. Atharv has over ten years of experience in applying ML solutions in healthcare, and primarily helping the commercial pharma industry realize the enhanced impact of leveraging RWD advanced analytics and ML-based techniques.

Arindam Paul is a Data Science Manager at ZS's Bangalore office, and a core member of the firm's RWD and AI COEs. He brings over a decade of experience in generating insights through AI and ML solutions, with a strong focus on commercial and medical applications in the pharmaceutical industry.

Prashant Mishra is a Data Science Consultant at ZS's New Delhi office, leading ML/AI based solution development efforts across commercial and medical functions, with primary focus on patient journey analytics using RWD. With over five years of experience, Prashant has delivered data-driven solutions to address a range of business challenges within oncology and rare disease therapeutic areas.

Arrvind Sunder is a principal from ZS's Evanston office and has been with ZS for 17 years. Arrvind is a leader in ZS's RWD and insights venture where he focuses his time on enabling clients to realize competitive differentiation through data, digital and AI.

Using Machine Learning to Establish the Impact of Patient Support Services and the Coordinated Efforts of Field Reimbursement and Access Specialists on Therapy Fulfillment Rates, Time to Fulfillment, and Patient Adherence

Prabhjot Singh, Senior Director, Axtria Inc.; Hemant Kumar, Director, Axtria Inc.; Aayush Gupta, Manager, Axtria Inc.

Abstract: Pharmaceutical companies invest heavily in programs like Field Reimbursement and Access Specialists (FRAS) and Nurse Navigators (NN) to improve patient outcomes. These programs are designed to provide comprehensive support to patients, ensuring they receive timely access to therapies and maintain adherence to prescribed treatments.[1] Despite significant investments, it has been challenging to quantify the effectiveness of these programs in improving patient outcomes.[2]

To address this need, Axtria engineered an advanced predictive analytics framework. This sophisticated machine learning (ML) system integrates data from various sources, including hub services, anonymized patient-level data (APLD), social determinants of health (SDOH), and field team activities. This integration allows Axtria to identify delays in medication fulfillment and predict non-fulfillment and non-adherence. By shifting from reactive to predictive strategies, the framework enhances both patient outcomes and operational efficiency.[3]

Axtria's models demonstrated strong predictive accuracy, with high precision/recall and low root mean squared error (RMSE). Early identification of delays and non-adherence risks enabled tailored interventions such as refill reminders, telehealth follow-ups, and financial assistance. This proactive approach not only improves patient adherence but also optimizes resource allocation and operational processes.

This study quantifies the impact of support programs on key patient metrics, including fulfillment rate and adherence. Axtria's findings highlight the potential of ML-powered predictions and integrated patient data to address challenges in managing specialty therapies. Leveraging these advanced analytics allows stakeholders to improve patient outcomes and the efficiency of healthcare delivery systems.

Introduction

Pharmaceutical companies deploy various patient support programs to ensure patients can start and continue their prescribed therapies effectively. These programs, each with distinct objectives, are crucial in addressing the diverse needs of patients. Key objectives of these programs include:

- **Providing support during claim approvals:** Assisting patients while they await insurance claim approvals to ensure timely access to medications. This support is vital as delays in claim approvals can lead to interruptions in therapy, negatively impacting patient health outcomes.[1]

- **Navigating insurance coverage and eligibility:** Helping patients understand and navigate the complexities of insurance coverage and eligibility requirements. This includes assistance with understanding policy details, coverage limits, and eligibility criteria, which can be particularly challenging for patients with complex or rare conditions.[2]

- **Supporting therapy adherence:** Offering continuous support to help patients remain on their prescribed therapies. This includes interventions such as medication reminders, counseling, and follow-up calls to ensure patients adhere to their treatment plans.[3]

Given the varied goals of these programs, it is essential to measure their success using specific patient success metrics. This model aims to:

1. **Assess the impact of patient support programs on treatment initiation, adherence, and treatment continuation:** Evaluating how well these programs achieve their intended outcomes. This involves analyzing data to determine whether patients are starting their therapies on time, adhering to their prescribed regimens, and continuing their treatments over the long term.

2. **Evaluate the effectiveness of individual and combined programs to identify synergies:** Understanding how different programs work together and their combined impact on key patient metrics. By examining the interactions between various support programs, this study identifies synergies that enhance overall effectiveness.

3. **Examine how programs influence payer types, pharmacies, and out-of-pocket costs:** Analyzing the broader effects of these programs on various aspects of the healthcare system. This includes understanding how different payer types (e.g., commercial insurance, Medicare) and pharmacy types (e.g., specialty pharmacies) are impacted by support programs and how these programs affect patients' out-of-pocket costs.

Through the analysis of key patient metrics such as fulfillment and adherence, this study provides a comprehensive evaluation of patient support programs and their effectiveness in improving patient outcomes. Focusing on an oncology drug prescribed for prostate cancer treatment, this evaluation will help pharmaceutical companies optimize their support programs, allocating resources effectively to maximize patient benefits

In summary, this study aims to provide a detailed assessment of patient support programs, leveraging advanced analytics to better understand their effect on therapy initiation, adherence, and continuation. By doing so, Axtria hopes to contribute to the development of more effective support strategies that improve patient outcomes and operational efficiency in the healthcare system.

Methods

The study employed a comprehensive approach, using multiple predictive models to evaluate patient behavior in therapy initiation and continuation.[4] The integration of diverse data sources and advanced modeling techniques enabled the identification of high-risk patients and quantified the impact of various patient support programs. A thorough lookahead period was also incorporated to account for impacts on patient health outcomes beyond the analysis period.[1]

Data Collection and Integration

The study's authors integrated data from multiple sources, including hub and longitudinal claims, to develop predictive features related to fulfillment and adherence.

This involved:

- **Specialty Pharmacy Claims Data:** Used to identify a relevant patient cohort and analyze refill patterns, out-of-pocket costs, and payer types. This data provided insights into patient behavior and financial barriers to therapy.[2]

- **Client Support Programs:** Data from programs such as Field Reimbursement Access Specialists (FRAS), Nurse Navigators (NN), and Copay Cards. These programs offer various forms of support to patients, and their effectiveness needed to be quantified.[3]

- **Patient Demographics:** Information such as age, gender, and other demographic factors that could influence patient behavior and therapy outcomes.

- **Social Determinants of Health (SDOH):** Factors such as economic stability, education level, social behavior, and ethnicity. These variables were crucial in understanding the broader context of patient health and access to care.

- **Formulary Access:** Data on whether patients had favorable or unfavorable formulary access, which affected their ability to obtain prescribed medications.

Feature Engineering

The study's authors defined business rules to calculate each outcome metric:

- **Fulfillment:** The first time the patient receives a successful paid shipment within the analysis period. This metric shows the initial success of starting therapy, focusing on patients who begin their treatment.[1]

- **Persistence:** The total days on therapy for a patient post-fulfillment within the analysis and look-forward periods. This metric measures how long patients continue their therapy.[2]

- **Adherence:** The rate at which a patient fills their prescriptions relative to the total length of therapy. This metric assesses adherence to the prescribed regimen and is measured using the Medication Possession Ratio (MPR).[3]

- **Time-to-Fulfillment:** The time elapsed between the first referral and the first shipment of the drug. This metric evaluates the efficiency of the therapy initiation process.

The authors used attribution methodology to assign calls deployed to address access barriers at the patient level. Patient-level calls were categorized as "proactive" or "reactive" based on the patient's interaction with detailing efforts. For example, calls to the healthcare provider (HCP) of a new-to-brand (NTB) patient within -90 days from the referral date were classified as "proactive," while calls within +90 days were classified as "reactive."

Direct mapping was available at the patient level for other programs, and binary flags were created to classify patients. Patient attributes like age, payer coverage, and socio-economic factors were incorporated to enhance the predictive power of the models.

Predictive Modelling

Once individual data tables were processed at the patient level, a master data table was created for modeling. The study was structured around four predictive models:[4]

- **Fulfillment Model:** Predicts whether a patient is likely to initiate non-free therapy based on their program interactions and individual profile. This model helps identify patients who may face barriers to starting their therapy.[1]

- **Persistence Model:** Assesses the likelihood of a patient continuing non-free therapy beyond a specific timeframe. This model is crucial for understanding long-term therapy adherence.[2]

- **Adherence Model:** Evaluates the consistency of patient adherence with a prescribed therapy. This model identifies patients at risk of non-adherence and helps target interventions.[3]

- **Time-to-Fulfillment (TTF) Model:** Estimates the days between referral and therapy initiation. This model helps optimize the process of getting patients started on their therapy.

To forecast fulfillment and persistence probabilities, XGBoost and random forest classification models were applied.[4] Machine learning-based regression and long short-term memory (LSTM)-based time-series models were utilized to monitor changes in adherence behavior over time. Hyperparameters were fine-tuned to maximize model accuracy. Evaluation metrics—including accuracy, precision, recall, R-squared, and RMSE—were analyzed to identify the superior model.

Impact Calculation and Response Curves

The impact of each support program was calculated by comparing the average historical outcome with the average predicted outcome when the program was excluded. Additionally, each feature's average SHAP value (Shapley Additive Explanations) was calculated across all instances to determine its overall impact on the model's predictions. SHAP values were used to quantify the impact of each feature on the predicted outcomes, helping to identify which features drive the changes in patient outcomes when support programs are implemented or varied.[5]

A sensitivity analysis was performed to understand the impact on each outcome metric if the reach and frequency of a program varied. SHAP dependence plots were created to visualize how changes in key features affect the predicted outcomes, providing insights into the model's sensitivity to variations in these features. This analysis offered valuable insights into how changes in program implementation could affect patient outcomes1.

This methodological framework allowed the quantification of the attributable impact of marketing tactics such as FRAS, NN programs, and patient support programs on key patient outcomes. Additional insights included program effectiveness across payer types, the impact of pharmacy-specific programs, and the effect of programs on patients facing high out-of-pocket costs. By understanding these impacts, support programs can be better tailored to meet the needs of different patient populations and improve overall therapy outcomes.[2]

Results

The study's results are presented based on the performance of various predictive models and their impact on patient outcomes. The authors evaluated each model using specific metrics to determine their accuracy and effectiveness.

Prediction Accuracy

All models achieved good accuracy, with precision values greater than 80% and RMSEs less than 0.2. Here are the detailed results for each model:

- **Fulfillment Model:** This model achieved an accuracy of 76%, with a precision of 82% and a recall of 81%. The RMSE was not applicable for this classification model.
- **Persistence Model:** This model showed an accuracy of 79%, with a precision of 73% and a recall of 89%. Similar to the fulfillment model, RMSE was not applicable.
- **Adherence Model:** The adherence model, being a regression model, had an RMSE of 0.12, indicating a good fit.
- **Time-to-Fulfillment Model:** This regression model had an RMSE of 0.19, demonstrating its predictive accuracy.

Meric	Fulfillment Model	Persistence Model	Adherence Model	Time to Fulfillment
Accuracy	76%	79%	not applicable	not applicable
Precision	82%	73%	not applicable	not applicable
Recall	81%	89%	not applicable	not applicable
RMSE	not applicable	not applicable	0.12	0.19

Table 1: Performance metrics across four models.

Receiver Operating Characteristic Area Under the Curve (ROC AUC) was generated to illustrate the performance of the classification model by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. The model achieved ROC = 0.85, demonstrating that the model has a good ability to distinguish between positive and negative classes, with higher values representing better performance.

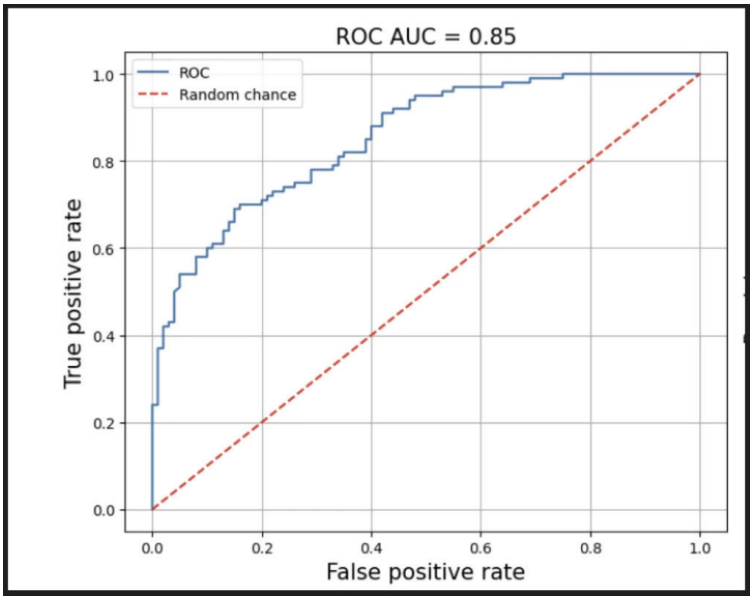


Figure 1: ROC curve analysis: true positive rate vs. false positive rate.

Key Predictors: The models identified payer type and the utilization of copay cards as the most influential predictors across all analyses. The top predictors specific to each model are outlined as follows:

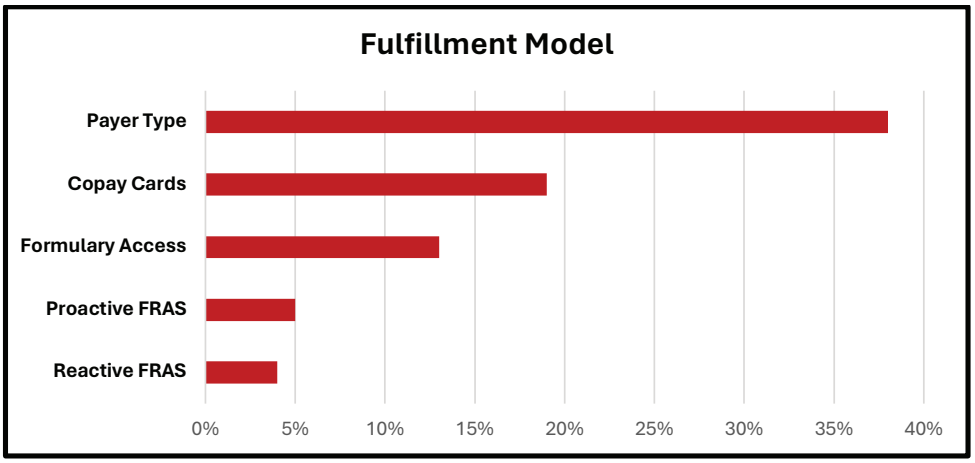


Figure 2a: Fulfillment model.

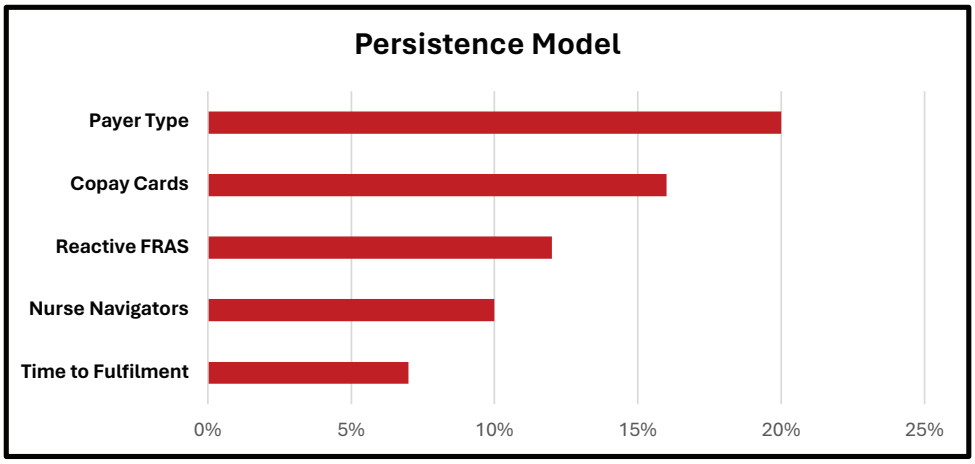


Figure 2b: Persistence model.

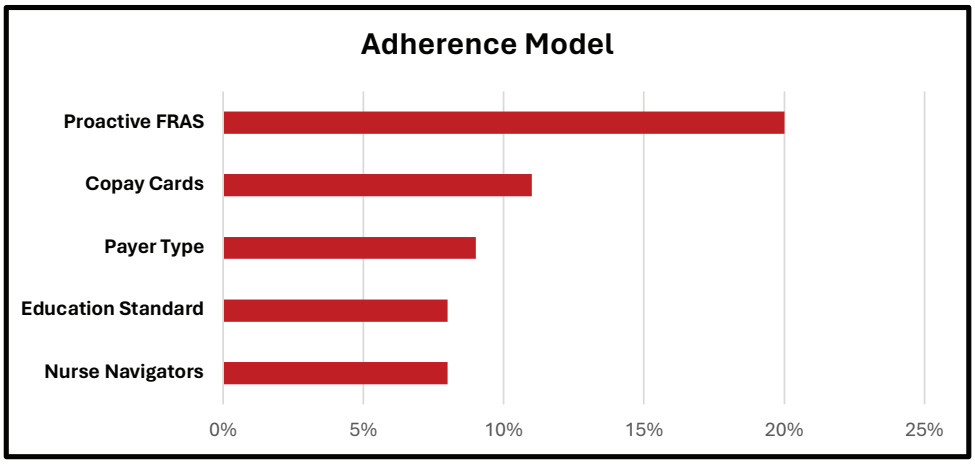


Figure 2c: Adherence model

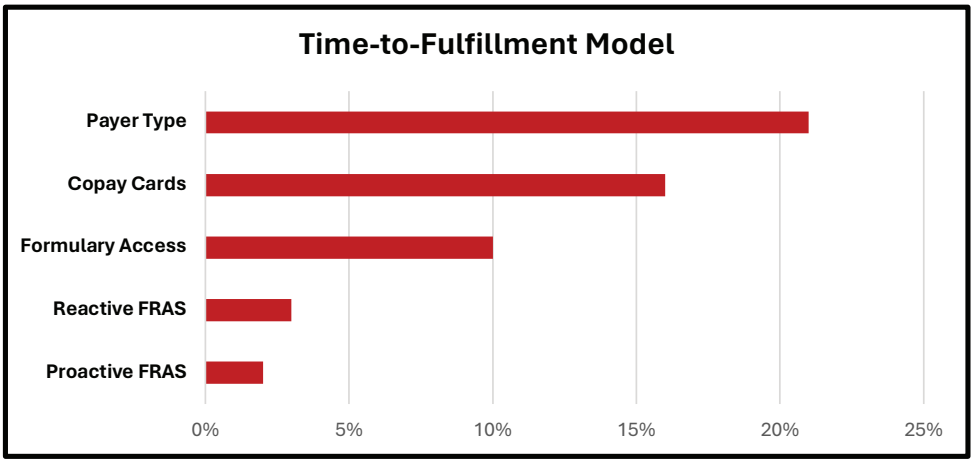


Figure 2d: Time-to-Fulfillment model.

Impact and Synergy Assessment: Support programs significantly improved overall patient outcomes from their baseline levels. Here are the detailed impacts:

- **Fulfillment Rate:** The baseline fulfillment rate was 59.0%. With the implementation of support programs, the incremental impact was +9.5%, resulting in an overall fulfillment rate of 68.5%.
- **Persistence Rate:** The baseline persistence rate was 54.8%. Support programs increased this by +7.5%, leading to an overall persistence rate of 62.3%.
- **Adherence:** The baseline adherence rate was 87.8%. Support programs improved this by +2.3%, resulting in an overall adherence rate of 90.1%.
- **Time-to-Fulfillment:** The baseline time to fulfillment was 19 days. Support programs reduced this by one day, resulting in an overall time to fulfillment of 18 days.

	Fulfillment Rate (in %)	Persistence Rate (in %)	Adherence (in %)	Time to Fulfillment (in days saved)
Baseline Impact	59.0%	54.8%	87.8%	19 days
Incremental Impact	+9.5%	+7.5%	+2.3%	-1 day
Overall Impact	68.5%	62.3%	90.1%	18 days

Table 2: Baseline and incremental impact assessment of each metric.

"Baseline impact" refers to the impact obtained without any support program engagements, while "overall impact" includes both baseline and incremental impacts. "Incremental impact" combines individual support program impacts and synergies.

Support programs increased overall patient fulfillment by 9.5%, with individual programs accounting for 6.8% and the remaining 2.7% due to synergistic effects. Proactive and reactive FRAS exposure compounded positive impacts on initiation. Combined efforts of FRAS and Nurse Navigators helped patients stay on therapy longer, and faster fulfillment led to longer therapy persistence.

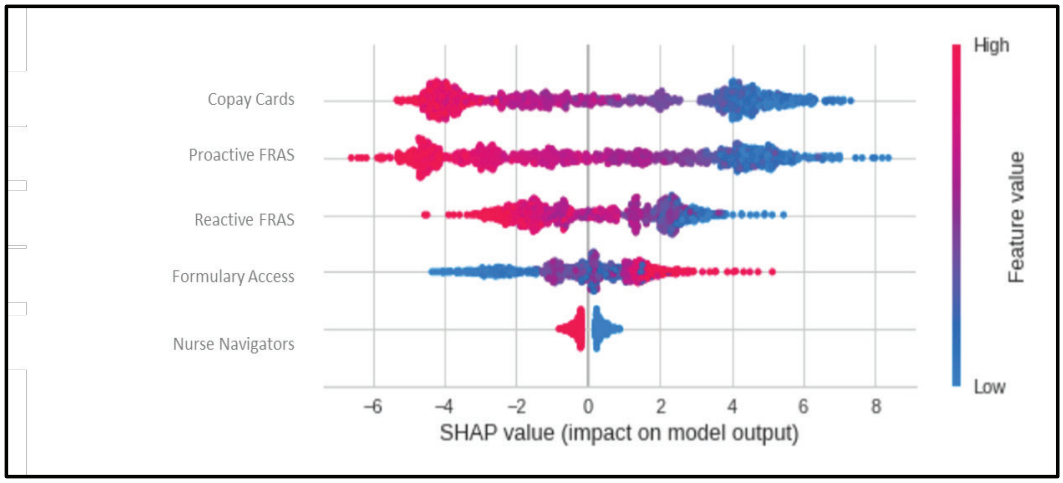


Figure 3: SHAP values for model variables.

Simulation: The simulation results showed how varying the frequency of FRAS interactions impacted fulfillment and persistence rates. Here are the detailed findings:

- Reducing FRAS by 50% decreased the fulfillment rate to 67.00% (a negative 1.60% impact from historical levels) and the persistence rate to 60.60% (a negative 1.70% impact). If FRAS is reduced by 20%, fulfillment and persistence rates fall to 68.30% and 61.80%, respectively.

- Increasing FRAS by 20% raised the fulfillment rate to 69.30% (a positive 0.70% impact from historical levels) and the persistence rate to 63.00% (a positive 0.80% impact). If FRAS is increased by 50%, fulfillment and persistence rates rise to 70.10% and 64.10%, respectively.

FRAS Frequency	Fulfillment Rate	% Impact (Incremental)	Persistence Rate	% Impact (Incremental)
Reduced by 50%	67.00%	-1.60%	60.60%	-1.70%
Reduced by 20%	68.30%	-0.20%	61.80%	-0.50%
Historical Activity Level	68.50%	-	62.30%	-
Increased by 20%	69.30%	0.70%	63.00%	0.80%
Increased by 50%	70.10%	1.60%	64.10%	1.80%

Table 3: Impact of FRAS frequency on fulfillment and persistence rates.

Discussion

The findings of this study demonstrate that patient support programs significantly enhance therapy initiation, adherence, and continuation. The predictive models developed by Axtria provided valuable insights into the factors influencing patient behavior and the effectiveness of various support programs.

Impact of Support Programs

The results indicate that support programs, particularly Field Reimbursement and Access Specialists (FRAS) and Nurse Navigators (NN), play a crucial role in improving patient outcomes. The proactive and reactive strategies employed by FRAS were particularly effective in increasing fulfillment rates and reducing time to therapy initiation. These strategies ensured patients received their medications promptly, minimizing delays that could negatively impact their health.

Nurse Navigators contributed significantly to patient persistence, ensuring that patients remained on therapy for longer periods. By providing continuous support and addressing patient concerns, Nurse Navigators helped maintain high levels of adherence and persistence, which are critical for achieving positive health outcomes.

Synergies and Combined Effects

The study also highlighted the synergistic effects of combining different support programs. For instance, the combined efforts of FRAS and NN resulted in higher persistence rates, demonstrating the importance of a coordinated approach to patient support. The interaction between proactive and reactive FRAS strategies compounded the positive impacts on therapy initiation, while the collaboration between FRAS and NN ensured sustained patient engagement and adherence.

These synergies underscore the value of integrating multiple support mechanisms to address the diverse needs of patients. By leveraging the strengths of different programs, pharmaceutical companies can create a more comprehensive support system that enhances overall patient outcomes.

Predictive Accuracy and Model Performance

The high accuracy and precision of the predictive models underscore the potential of machine learning in healthcare. By accurately predicting patient behavior, these models enable targeted interventions, improving both patient outcomes and operational efficiency. The models' strong performance metrics, such as high precision and low RMSE, indicate their reliability in identifying high-risk patients and anticipating therapy adherence and persistence.

The use of advanced machine learning techniques, such as XGBoost, random forest, and LSTM-based time-series models, allowed for the development of robust predictive models. These models can be continuously refined and updated with new data, ensuring their ongoing relevance and accuracy.

Program Effectiveness Across Different Patient Segments

The analysis revealed that the effectiveness of support programs varied across different patient segments. For example, copay cards were most effective for patients with commercial coverage, while FRAS had a more significant impact on patients with Medicare coverage. These insights can inform the design of tailored support strategies to address the specific needs of different patient groups.

Understanding the differential impact of support programs across various patient demographics and payer types is crucial for

optimizing resource allocation. By targeting interventions based on patient-specific factors, pharmaceutical companies can enhance the effectiveness of their support programs and improve patient outcomes.

Limitations and Future Research

While the study provides valuable insights, it is not without limitations. The models were developed using data from specific programs and patient cohorts, which may limit their generalizability. Future research should explore the application of these models to other therapeutic areas and patient populations to validate and extend the findings.

Additionally, the study focused on a limited set of support programs and patient metrics. Expanding the scope to include other types of support programs and additional patient outcomes could provide a more comprehensive understanding of the impact of patient support initiatives.[6]

Future research should also investigate the long-term effects of support programs on patient health outcomes and healthcare costs. By examining the sustained impact of these programs, researchers can provide more robust evidence to support their continued implementation and optimization.

Conclusion

This study demonstrates the significant impact of patient support programs on therapy initiation, adherence, and continuation. By leveraging machine learning models, the research team was able to identify high-risk patients, predict their behavior, and quantify the effectiveness of various support programs. The findings underscore the importance of a coordinated approach in patient support, highlighting the synergistic effects of combining different programs.

Integrating data from multiple sources—including hub services, anonymized patient-level data (APLD), social determinants of health (SDOH), and field team activities—allowed for a comprehensive analysis of patient behavior. The predictive models developed in this study, including the Fulfillment, Persistence, Adherence, and Time-to-Fulfillment models, demonstrated high accuracy and precision, enabling targeted interventions that improve patient outcomes and operational efficiency.

The study revealed that support programs such as Field Reimbursement and Access Specialists (FRAS) and Nurse Navigators (NN) play a crucial role in enhancing patient outcomes. The proactive and reactive strategies employed by FRAS were particularly effective in increasing fulfillment rates and reducing time to therapy initiation. Nurse Navigators contributed significantly to patient persistence, ensuring that patients remained on therapy for longer periods.

This retrospective study analyzed the impact of various patient support programs on therapy adherence and its subsequent effect on health outcomes. The findings indicate that these programs played a critical role in helping patients initiate therapy and adhere to prescribed doses, which are essential for achieving the intended therapeutic benefits. By facilitating consistent medication adherence, support programs contribute to improved overall patient health, reinforcing their value in healthcare interventions.

The synergistic effects of combining different support programs were also highlighted, with the combined efforts of FRAS and NN resulting in higher persistence rates. This result demonstrates the importance of a coordinated approach in patient support, where multiple programs work together to address the diverse needs of patients.

The analysis also revealed that the effectiveness of support programs varied across different patient segments. For example, copay cards were most effective for patients with commercial coverage, while FRAS had a greater impact on patients with Medicare coverage. These insights can inform the design of tailored support strategies to address the specific needs of different patient groups.

While the study provides valuable insights, it is not without limitations. The models were developed using data from specific programs and patient cohorts, which may limit their broader use. Future research should explore the application of these models to other therapeutic areas and patient populations. Expanding the scope to include different types of support programs and patient outcomes may provide a more comprehensive understanding of the impact of patient support initiatives.

In conclusion, the insights gained from this study can inform the design of more effective patient support strategies, ultimately improving patient outcomes and operational efficiency in managing specialty therapies. By leveraging advanced analytics and machine learning, pharmaceutical companies can enhance their support programs, ensuring that resources are allocated effectively to maximize patient benefits.

References

1. 3 predictions for patient support programs in 2025. AllazoHealth. December 18, 2024. Accessed January 14, 2025. https://allazohealth.com/resources/3-predictions-for-patient-support-programs/

2. Lenz F, Harms L. The impact of patient support programs on adherence to disease-modifying therapies of patients with relapsing-remitting multiple sclerosis in Germany: a non-interventional, prospective study. *Adv Ther.* 2020 Jun;37(6):2999-3009. doi: 10.1007/s12325-020-01349-3

3. Rubin D, Mittal M, Davis M, Johnson S, Chao J, Skup M. Impact of a patient support program on patient adherence to adalimumab and direct medical costs in Crohn's Disease, ulcerative colitis, rheumatoid arthritis, psoriasis, psoriatic arthritis, and ankylosing spondylitis. *J Manag Care Spec Pharm.* 2017; 23(8). https://www.jmcp.org/doi/10.18553/jmcp.2017.16272

4. Molnar C. *Interpretable machine learning: a guide for making black box models explainable.* Independently published; 2022.

5. Apley DW, Zhu J. Visualizing the effects of predictor variables in black box supervised learning models. *J R Stat Soc Series B Stat Methodol.* 2020; 82(4):1059-1086. https://doi.org/10.1111/rssb.12377

6. Wills A, Mitha A, Cheung WY. Data collection within patient support programs in Canada and implications for real-world evidence generation: the authors' perspective. *J Pharm Pharm Sci.* 2023 Oct 13;26:11877. doi: 10.3389/jpps.2023.11877

Acknowledgments

This research was supported by Axtria Inc. The authors would like to express their gratitude to the administrative staff at Axtria Inc. for their invaluable support throughout this project. Special thanks are extended to the IT department for its technical assistance, which was crucial for the successful completion of this study. Additionally, the authors are deeply grateful to their families for their unwavering support and understanding during the research process.

About the Authors

Prabhjot Singh is a senior director at Axtria. Over 17 years, he has served clients globally in the pharma, media, and technology sectors. Combining data with innovative techniques, his expertise cuts across impact measurement, launch tracking, and sales force sizing.

Hemant Kumar is a director at Axtria. Partnering with global life sciences organizations, he delivers impactful solutions in marketing, sales, and patient analytics. His machine learning initiatives demonstrate real-world success in areas such as segmentation and promotional effectiveness.

Aayush Gupta is a manager at Axtria with a repertoire of commercially successful projects in marketing mix modeling, patient analytics, and promotional measurement.

Enhance your insights by leveraging public domain data sources

JP Tsang, PhD and MBA (INSEAD), President of Bayser Consulting;
Shunmugam Mohan, VP of Operations, Bayser Consulting;
Maylis Larroque, Senior Consultant, Bayser Consulting;
Zhangjin Xu, PhD, Senior Consultant, Bayser Consulting

Abstract: There is a plethora of high-quality data sources in the public domain, meaning they are absolutely free of charge. They can greatly enhance the insights we glean from commercial data sources in at least three ways. First, they can help expose the holes, biases, and anomalies present in commercial data sources. Second, they can offer perspectives that commercial data sources simply cannot. Third, they can provide directional answers in the absence of commercial data, which may largely suffice for an initial exploration of the question.

Yet, most data analysts do not use these data sources. Interestingly, they do not avoid these data sources knowingly due to past bad experiences. They are simply unaware of their existence. Evidence? Anecdotal probing of scores of analysts indicates this is the case.

How can that be? Economists tell us that the lower the price of a good or service, the higher the demand. However, economists implicitly assume that the consumer is aware of the product and knows that it is free. This is simply not true. We call this the Paradox of the Free. Because the data is free, there are no dog-and-pony shows to sing its praises, no email solicitations to entice analysts to get started, and no one to answer data questions via phone or chat. Furthermore, from the analyst's standpoint, it is more convenient to use commercial data sources since they come with customer support. Public domain data sources, on the other hand, offer no such support. Needless to say, the analyst does not personally benefit from any financial savings by using free public data sources instead of commercial ones.

In this paper, we'll discuss use cases that pervade virtually all therapeutic areas. For each of them, we'll specify the business question and identify which free data sources are useful for answering the question or enhancing the answers provided by commercial data sources. Below are the use cases we'll focus on:

1. Identification of Physicians for Segmentation and Targeting
2. Identification of Hospitals for Segmentation and Targeting
3. Mapping of the Payer Landscape for Contracting
4. Development of Look-Alike Models for Expanding Therapy Usage

Note on Data Sources

Public domain data sources come in different shapes and sizes. Two dimensions stand out: (1) relevance and (2) ease of access. In several instances, we may not recommend a data source in the public domain for any use case, not because it is irrelevant or only marginally relevant, but because access to the data is a real challenge. Below is a simple classification we use.

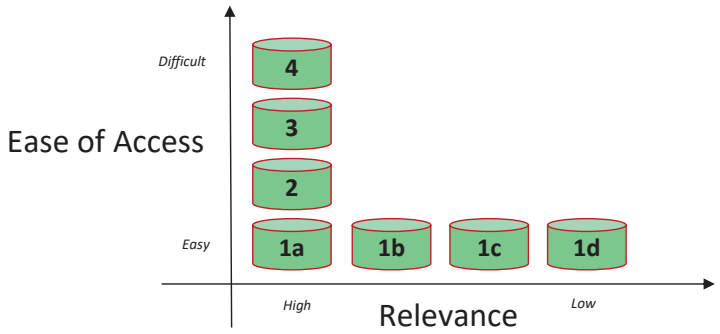


Fig 1 – Ease of Use and Relevance of Data Sources

1. No problem with the download, and
    a. The data fits the bill perfectly. Example: NPPES (National Plan and Provider Enumeration System) lists all the providers in the country.
    b. The data is dated. Example: Medicare Part B and Part D are a couple of years old.
    c. The data lacks granularity. For example, we want the data at the county level but it is at the state level. SHADAC (State Health Access Data Assistance Center) provides excellent insurance status data at the state level, not at the county level or lower.
    d. The identifier used is not universal. For instance, instead of using the NPI to reference a physician, the data uses an internal id. Example: NPDB (National Practitioner Data Bank) is a confidential data source regarding disciplinary actions and malpractice payments of healthcare practitioners. The ID of the physician is encrypted.
2. There are hoops to jump through to get the data. For instance, a motivation letter needs to be submitted and a committee will decide if access is to be granted or not. Another scenario is promise in writing to publish the findings of the analysis. Example: MedPAR which is an LDS (Limited Data Set) from CMS, requires a motivation letter and there is a fee if access is granted.
3. Downloading data is a real challenge. A prime example is TIC (Transparency in Care) data. That's because the data is humongous and is scattered over tens of thousands of files and hundreds of sites. Another example is ZocDoc, which provides detailed information on the physician and the list of insurance companies they accept. The data is served up from a dashboard and may not be scraped easily from the web.
4. We do not meet the eligibility criteria to have access to the data. For example, RIF (Research Identifiable Files) files from CMS can only be accessed by academic researchers in universities.

There are also commercial data sources that are outside the scope of this paper. For IDNs (Integrated Delivery Networks), for example, Definitive Healthcare and IQVIA OneKey, which include HCOS (Health Care Organizations and Systems), are good data sources.

Use Case 1 - Identification of Physicians for Segmentation and Targeting

Objective: Identify the most relevant physicians to target to increase drug prescriptions

Approach: Construct a multi-faceted picture of the physicians by looking at the following:

1. Profile of the physician and their affiliation with group practices and hospitals
2. Drug Prescription Activity to capture the brands and volume of drugs prescribed
3. Patient Referral Activity to map out the connections between physicians
4. Clinical Trial Activity to identify the clinical trials a physician conducts
5. Manufacturer Networking to understand the financial relationships between physicians and manufacturers

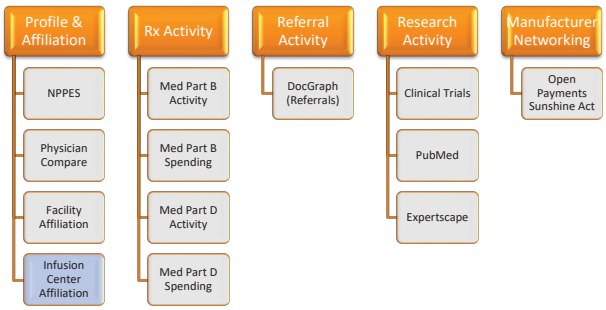


Fig 2 – List of Data Sources in the Public Domain for Segmenting and Targeting Physicians

A. Profile and Affiliation

NPPES (National Plan and Provider Enumeration System) – This data source lists all 7 million providers: 5 million individual providers including physicians, NPs/PAs, nurse practitioners, therapists and other licensed practitioners, and 2 million institutional providers including hospitals, skilled nursing facilities, ambulatory care facilities, pharmacies, home health agencies and hospices, labs and the like. NPPES indicates, for each provider, the name, address, city, state, and ZIP among other things which allows us to get a count of the number of providers by specialty for each county. It is managed by the Centers for Medicare & Medicaid Services (CMS), a division of the U.S. Department of Health and Human Services and is updated daily.

Physician Compare – This data source provides detailed profile information on approximately one million physicians who participate in Medicare and is managed by CMS (Centers for Medicare & Medicaid Services). In addition to basic demographics and quality measures such as patient experience, the data source indicates the group practice(s) the physician is affiliated with. Note the group practice is one of several entities the physician is affiliated with. The physician is identified through an NPI (National Provider ID) and the group practice through an ORG_PACID (Organization PECOS Associate Control ID where PECOS stands for Provider Enrollment, Chain, and Ownership System). The dataset is updated monthly. The most recent version, as of the writing of this article, is March 7, 2025.

Facility Affiliation – This data source is contained within a larger data source called PECOS (Provider Enrollment, Chain, and Ownership System). PECOS contains enrollment and affiliation information for a wide range of healthcare providers, including hospitals, individual physicians, ambulatory surgical centers, critical access hospitals,

skilled nursing facilities, home health agencies, hospices, clinical laboratories, and durable medical equipment suppliers. The data source includes information on about 6,000 hospitals and identifies the physicians through an NPI (National Provider Id) and the hospital using a CCN (CMS Certification Number). The PECOS system is updated continuously, so the most recent data is generally only a few days to a couple of weeks old.

Infusion/Dialysis Center Affiliation – Affiliation is definitely a misnomer, as physicians practicing in their offices are not affiliated with any such centers. What this data portrays is the connection between care centers that patients go to and physicians. This "affiliation" data comes in handy to measure the activity of physicians when the patients of the physician get their drugs not at the physician's office but in an infusion center or dialysis center. This data source is not available off-the-shelf in the public domain but rather can be constructed using paid PLD (Patient Level Data) data sources by tracking providers that patients visit in succession.

B. Prescription Activity

Medicare Part B Activity – This data source captures injectable and infused drugs administered by physicians, as well as certain oral drugs related to conditions like ESRD (End-Stage Renal Disease) and preventive vaccines. The setting includes physician offices, hospital outpatient departments, and even patients' homes in some cases. This data reports the activity of individual physicians in terms of HCPCS (Healthcare Common Procedure Coding System) codes and to each HCPCS code is associated: (1) number of claims (about 10 million records), (2) average amount submitted, (3) average amount allowed, and (4) average amount Medicare paid. The activity of physicians with 10 beneficiaries or fewer is suppressed. This data source is managed by CMS (Centers for Medicare & Medicaid Services) and pertains to only FFS (Fee-For-Service) or Original Medicare portion only. It does not contain the Managed Medicare portion. The data is 2 years old.

Medicare Part B Spending by Drug – This dataset is the companion to the Medicare Part B Activity file in that it also reports activity by HCPCS code but aggregated at the national level. The "10 or fewer beneficiaries" suppression rule is still in force, but it rarely applies since the data is aggregated at the national level. This is why this data source comes in handy to indicate how much the Medicare Part B Activity file leaves out because of the suppression rule.

Medicare Part D Activity – This data source captures prescription drug coverage for medications primarily self-administered by beneficiaries. It reports the activity of individual physicians at the branded-generic drug name level, and for each branded-generic name, the following are provided: (1) number of claims, (2) 30-day fills, (3) days supply, (4) drug cost, and (5) number of beneficiaries. The activity of physicians with 10 or fewer beneficiaries is suppressed. This data source is managed by CMS (Centers for Medicare & Medicaid Services) and captures both the stand-alone PDP (Prescription Drug Plan) portion and MAPD (Medicare Advantage Prescription Drug Plan) portion (Part C), which bundles medical and prescription benefits offered by private insurance, is not captured. The data is 1–2 years old.

Medicare Part D Spending by Drug – This dataset is the companion file to the Medicare Part D Activity file, as it also reports activity by branded-generic name but aggregated at the national level. The "10 or fewer beneficiaries" suppression rule is still in effect but rarely applies since the data is aggregated at the national level. This is why this data source is useful in estimating how much data the Medicare Part D Activity file omits due to the suppression rule.

C. Referral Activity

DocGraph is a data source from Medicare that indicates the number of patients who move from one provider (from-provider) to another provider (to-provider) by tracking successive Part D payments for Medicare patients. A few things to note. First, providers are not only physicians but can also be pharmacies, for instance, which may complicate some analyses. Second, patient movements pertain to the entire Medicare Part D population and, as such, do not break down the number of patients by diagnosis or drug. This is not necessarily a drawback, as the data source captures deep relationships that exist between providers. CareSet, a data vendor based in Houston, Texas, sells not only DocGraph data from CMS but also consultative services around MrPUP, allowing patients to be segmented by diagnosis and drug. Older versions of DocGraph are inexpensive, and some older versions are free. Third, what is captured is patient movements, not referrals, which, with the right business rules in place, serve as a good approximation for referrals.

D. Clinical Trial Activity

ClinicalTrials.gov – This is a publicly accessible registry and results database maintained by the U.S. National Library of Medicine (NLM) at the National Institutes of Health (NIH), which provides detailed information on clinical studies conducted worldwide, including interventional and observational studies on drugs, devices, behavioral interventions, and other therapies. This data source is useful as it specifies the identity of the principal investigator (PI) associated with each study, the locations of the sites where the study is conducted, the status of the study, and a description of the study, which allows for further profiling of the PI. The data dates back to the early 2000s. Note that it does not report the NPI (National Provider ID) of the PI.

PubMed – A free resource developed and maintained by the United States National Library of Medicine (NLM) at the National Institutes of Health (NIH). It offers access to a vast repository of biomedical and life sciences literature and contains over 35 million citations and abstracts. It is updated daily. In addition to the title and abstract of each article, the database provides information on the authors and their affiliations. PubMed is user-friendly, offering advanced search features, including filters for publication date, text availability, article type, and language. It also provides links to full-text content when available, either through PubMed Central or publisher websites, facilitating seamless access to complete articles.

Expertscape – This is an online platform designed to identify and rank medical expertise across more than 29,000 biomedical topics by analyzing scientific publications in PubMed. It scores each article based on the type of publication and journal prominence. The contributions of individual authors are evaluated by considering factors such as authorship position (e.g., first author, senior author) and the number of publications on the topic.

E. Manufacturer Networking

Open Payments from Sunshine Act – This is a publicly accessible database created under the Affordable Care Act's Sunshine Act, which mandates the reporting of financial relationships between healthcare providers and the pharmaceutical and medical device industries. The database was developed by CMS and was first launched in 2014. It contains detailed information on payments and transfers of value made to over 1 million physicians and teaching hospitals, aiming to promote transparency in financial interactions within the healthcare industry. This data source indicates very clearly how much money a physician receives from which pharmaceutical

company and for what drug. All the payments are dated. This data source has multiple uses. First, it helps identify physicians that the pharmaceutical industry perceives as most valuable. Second, it helps assess the allegiance of a physician to a company by looking at other payments the physician receives from different pharmaceutical companies. Third, it helps identify which pharmaceutical companies send representatives to which physicians. The giveaway here is lunch reimbursement.

Use Case 2 - Identification of Hospitals for Segmentation and Targeting

Objective: Identify hospitals to expand usage of our drug

Approach: Gather as much information as possible on individual hospitals, their activity, and the geographies in which they are located.

1. Profile – Key demographics of the hospital
2. System – Larger entity to which the hospital belongs
3. Big Picture – Broad-brush statistics regarding different types of hospitals or admissions at sub-national levels
4. Assessment – Rating and rank-ordering of hospitals based on multiple criteria
5. Catchment – Residential zip codes from which patients come and to which they go
6. Activity - Admissions and types of care individual hospitals provide

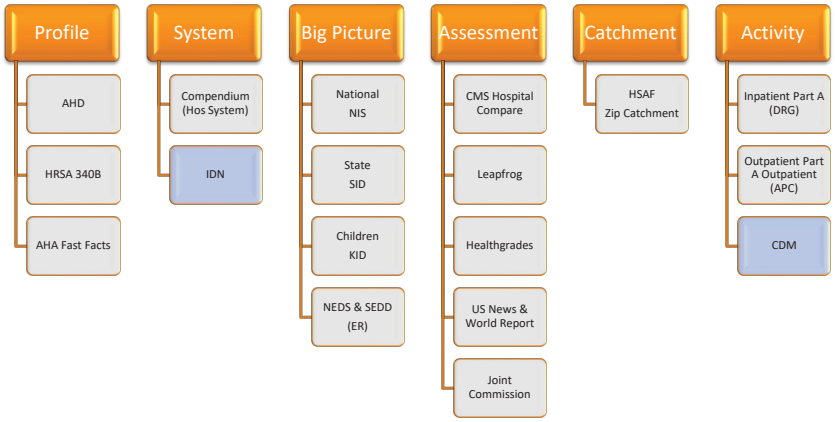


Fig 3 – List of Data Sources in the Public Domain for Segmenting and Targeting Hospitals

A. Hospital Data Sources – Profile

AHD (American Hospital Directory) – This data source provides the name and address of the institution, operational status (e.g., operational, closed, under renovation), type of facility (e.g., hospital, clinic, nursing home, rehabilitation center, urgent care center), control type (e.g., government, non-profit, for-profit), total staffed beds, total discharges, average daily census (average number of inpatients present), and average length of stay. AHD has data on 7,000+ hospitals, and the data is about 1 year old. The data is managed by American Hospital Directory, Inc., an independent organization that has no connection with the AHA (American Hospital Association) and comes from both public and proprietary sources. Public data sources include CMS (Centers for Medicare & Medicaid Services), MedPAR (Medicare Provider Analysis and Review), and Medicare cost reports.

HRSA 340B Covered Entity – This database indicates if the entity type of the hospital is DSH (Disproportionate Share Hospital), which means the hospital serves low-income patients and receives federal funding (about 1,300 hospitals are DSH), and also if the hospital participates in the 340B program, which requires manufacturers to provide outpatient medications at significantly reduced prices (about 1,600 hospitals participate in 340B). The HRSA (Health Resources and Services Administration) data source refers to the hospital by both a 340B identifier and an NPI along with the name and address of the hospital. The data is released yearly, 8 months into the following year, making the data 8–20 months old.

Fast Facts on US Hospitals from the AHA (American Hospital Association) describes key statistics, including the total number of hospitals (6,093), their classifications (e.g., 5,112 community hospitals), staffed beds (913,316), and total admissions (34,426,650 in 2013). More detailed information is available in the paid versions of data sources such as the AHA Guide, AHA Hospital Statistics, and the AHA Annual Survey Database, which provide comprehensive information on hospital demographics, organizational structures, service lines, utilization, finances, and staffing.

B. Hospital Data Sources – System

Compendium of U.S. Health Systems – This data source provides detailed information on health systems. As of 2023, the Compendium identifies 639 U.S. health systems and includes data on various components of health systems, such as hospitals, group practices, outpatient sites, nursing homes, and home health organizations. It is managed by the Agency for Healthcare Research and Quality (AHRQ), a division of the U.S. Department of Health and Human Services (HHS).

Let's quickly mention that there is a plethora of good data sources for IDNs (Integrated Delivery Networks) that are not in the public domain but are available for a fee: (1) HospitalView from Definitive Healthcare, (2) Annual Survey Database from AHA, (3) Healthbase Healthcare Affiliations Intelligence from Clarivate, and (4) OneKey Reference Data from IQVIA (includes HCOS – Health Care Organizations and Systems).

C. Hospital Data Sources – Big Picture

NIS (National Inpatient Sample) – This is a comprehensive, all-payer inpatient healthcare database developed as part of the Healthcare Cost and Utilization Project (HCUP) by the Agency for Healthcare Research and Quality (AHRQ). The data is available from 1988 through 2022 and is sampled from the SID (State Inpatient Database), including all

inpatient data that goes to HCUP. The NIS data captures 7 million patient stays from 4,500 hospitals, which are then projected to 35 million patient stays nationwide. The stays are broken down by patient age, gender, race, median income of ZIP code, primary and secondary diagnoses, procedures performed, length of stay, discharge status, total charges, payer, and the like. NIS also includes a separate file that provides information on each hospital, including bed size, teaching status, geographic location, and type of control. The ID of the hospital is encrypted as of 2012.

SID (State Inpatient Databases) – This is a set of all-payer, state-specific hospital inpatient databases developed as part of HCUP (Healthcare Cost and Utilization Project) by AHRQ (Agency for Healthcare Research and Quality). The data source captures inpatient discharge records from community hospitals in participating states. The data is broken down by patient sex, age, and, for some states, race. It also includes principal and secondary diagnoses and procedures, total charges, length of stay, expected primary payer (e.g., Medicare, Medicaid, private insurance, self-pay), admission source, and discharge status. Worth noting is that SID can be linked to hospital-level data from the AHA's (American Hospital Association) Annual Survey of Hospitals.

KID (Kids' Inpatient Database) – This is a comprehensive, all-payer pediatric inpatient care database developed as part of HCUP (Healthcare Cost and Utilization Project) by AHRQ (Agency for Healthcare Research and Quality). Patients are younger than 21 years of age. KID is released every three years, with data available from 1997 through 2019. Unweighted, it contains data from approximately 3 million pediatric discharges each year; when weighted, it estimates roughly 6 million hospitalizations. The KID includes a sample of 10% of normal newborns and 80% of other pediatric discharges from all U.S. community hospitals. The data is broken down by patient demographics (e.g., age, gender), clinical information (e.g., primary and secondary diagnoses, procedures performed), resource utilization (e.g., length of stay, total charges), and discharge status. Additionally, the KID provides a hospital file containing information on hospital characteristics such as bed size, teaching status, geographic location, and ownership type. The hospital identifiers are reassigned with each release, preventing the tracking of hospitals across different years.

NEDS (Nationwide Emergency Department Sample) and SEDD (State Emergency Department Databases) – These are both Emergency Department datasets, one at the national level and the other at the state level. NEDS is the largest all-payer ED database in the United States, with about 32 million ED visits annually, which is projected to about 137 million ED visits nationwide. The data contain information on patient demographics, visit characteristics, clinical diagnoses and procedures, and discharge dispositions. SEDD comprises state-specific databases that capture discharge information on all ED visits that do not result in hospitalization. The data contain details on patient demographics, visit reasons, clinical services provided, and discharge statuses.

D. Hospital Data Sources – Assessment

Hospital Compare – This data source from CMS evaluates all 4,500 or so hospitals that participate in Medicare using a star rating system (1 to 5 stars). The evaluation is based on clinical outcomes such as mortality and complication rates, readmission rates, patient safety measures like hospital-acquired infections and adverse events, and patient experience as measured by the HCAHPS (Hospital Consumer Assessment of Healthcare

Providers and Systems) survey. HCAHPS reports patient feedback on key aspects of hospital care—such as communication, staff responsiveness, cleanliness, pain management, and discharge instructions.

Leapfrog – This data source evaluates about 1,600 acute care hospitals out of the 6,000 or so hospitals in the country on measures like infection control, ICU staffing, infection prevention protocols, and similar factors, and assigns each hospital a letter grade from A to F. The data source is managed by the Leapfrog Group, a nonprofit organization in Washington, D.C. The data is available at the individual hospital level and is free of charge.

Healthgrades – This is a consumer-focused online platform maintained by Healthgrades, a privately held company based in Denver, Colorado. It provides profiles and star ratings (1 to 5) for about 4,000 hospitals and over 1 million physicians. Hospitals are assessed based on clinical outcomes (such as mortality and complication rates), patient safety indicators (including infection and readmission rates), and patient satisfaction survey results. Physicians are assessed based on board certifications, clinical outcomes in relevant specialties, patient reviews, and satisfaction scores.

Best Hospitals Ranking – This data source is from U.S. News & World Report, a reputable media organization based in Washington, D.C., and ranks 4,500 hospitals based on multiple criteria, including clinical outcomes, patient safety, nurse staffing, bed count, teaching status, availability of advanced technology, procedure volume, and reputation surveys among clinicians. The ranking is updated annually and has a one-year lag.

Joint Commission Accreditation – This data source is maintained by the Joint Commission, an independent nonprofit organization based in Oakbrook Terrace, Illinois, which accredits and certifies healthcare organizations such as acute care hospitals, critical access hospitals, ambulatory care centers, laboratory services, and specialty programs like cancer care and palliative initiatives. The database provides accreditation status and quality performance data for 6,200+ accredited hospitals and is accessible through the Quality Check portal.

E. Hospital Data Sources – Catchment

HSAF (Hospital Service Area File) – This data source indicates the number of patients admitted to a hospital and the zip code of origin of these patients for 4,500 hospitals. This data source is developed and maintained by the Dartmouth Atlas of Health Care, which is part of the Dartmouth Institute for Health Policy and Clinical Practice. Dartmouth uses raw data from CMS (Centers for Medicare & Medicaid Services) to build the data source. Note that HSAF does not break down the admission data by patient profile or type of service rendered in the hospital (diagnosis or procedure). The lag is 2–3 years.

F. Hospital Data Sources – Activity

Medicare Part A Inpatient by Provider – This data source provides information on inpatient discharges for Original Medicare Part A beneficiaries at the individual hospital level. It reports the number of discharges, average Medicare payments, and total payments by hospital. This data allows us to assess the inpatient Medicare activity of approximately 3,000 individual hospitals. Note that there is a more granular version, called Original Medicare Part A Inpatient by Geography and Service, which aggregates the same data at the MS-DRG (Medicare Severity Diagnosis Related Groups) and ZIP3 geographic levels. Both datasets are 8 to 12 months old.

Medicare Part B Outpatient by Provider – This data source provides information on outpatient services for Original Medicare Part B beneficiaries at the individual hospital level. It reports the number of services, average Medicare payments, and total payments by hospital. This data allows us to assess the outpatient Medicare activity of individual hospitals. Note that there is a more granular version of a related dataset, named Original Medicare Part B Outpatient by Geography and Service, which aggregates the same data at the Ambulatory Payment Classification (APC) and ZIP3 geographic levels. Both datasets are typically 8 to 12 months old.

CDM (Charge Data Master), a.k.a., Charge Master – This is an important type of data that captures the care a hospital provides to patients at the individual patient level. The CDM describes the medical procedures, diagnostic tests, and medications the patient receives day by day as an inpatient of the hospital. It is the source of the itemized bills for the patient and claims for the insurance payer. The identity of the hospital is not revealed. As far as we know, there is no CDM in the public domain. The paid ones are offered by data vendors such as Premier, IQVIA, and Health Verity.

Use Case 3 - Mapping of the Payer Landscape for Contracting

Objective: Map out the payer landscape at a granular geographic level to assess the merit of contracting with a specific payer.

Approach: Establish, at the county level (there are 3,144 counties in the country), which payers are present, how significant they are in the county, and the terms of the formulary in force regarding the drug of interest for these payers.

1. Payer Contracts to identify the Payers and their involvement in contracts
2. Drug Usage to assess Medicare and Medicaid activity
3. Insurance Status to shed light on the size of the insured and uninsured population

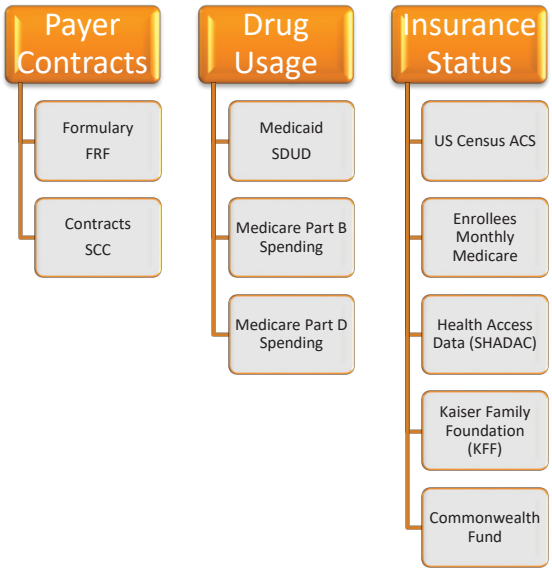


Fig 4 – List of Data Sources in the Public Domain for Mapping the Payer Landscape for Contracting

A. Payer Contracts

SCC (State County Contracts) – This data source lists all the commercial plans Medicare contracts with, both on the PDP side (patients who opt for Original Medicare) and the MAPD side (patients who opt for Managed Medicare), for each drug available under Medicare. Moreover, the data source indicates the name of the plan and the number of lives associated with each plan, giving us an idea of the relative significance of each plan for the market of interest.

FRF (Formulary Reference File) – This data source is more of a companion file, zeroing in on the plan and indicating the premium and the deductible. It also specifies the terms of the

formulary and whether any of the following apply: QL (Quantity Limited), ST (Step Therapy), PA (Prior Authorization), as well as the tier of the drug.

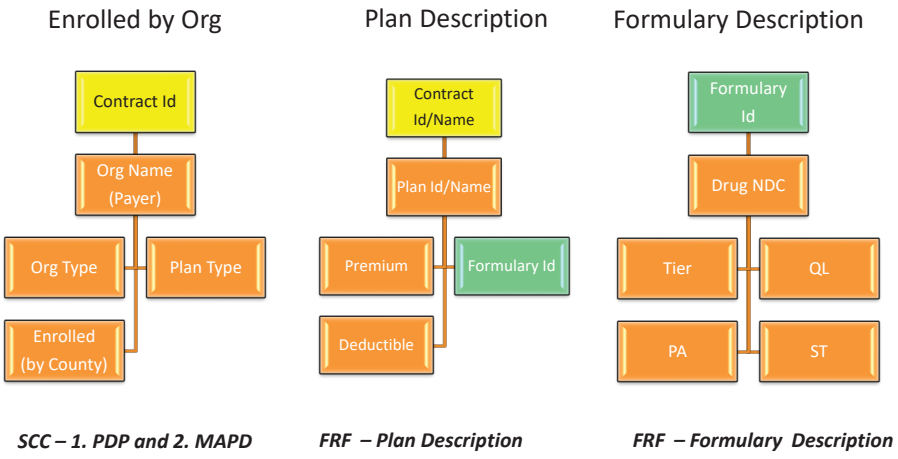


Fig 5 – Relationship between key variables in SCC and FRF

B. Drug Usage

Medicaid State Drug Utilization Data (SDUD) – This data source indicates Medicaid drug usage at the state level, not at the individual physician level. The drug is identified by its NDC code, and in some cases, there may be a large number of NDC codes for just one drug. The data is broken down by quarter. Drug usage is captured by the number of prescriptions and the dollar amount reimbursed—both by Medicaid and non-Medicaid sources. A flag indicates if suppression has been applied when too few patients were involved. The data source separates drug usage between Fee-for-Service Medicaid and Managed Care Medicaid.

Medicare Part B Spending by Drug – This dataset is the companion to the Medicare Part B Activity file, as it also reports activity by HCPCS code but aggregated at the national level. The "10 or fewer beneficiaries" suppression rule is still in force, but it rarely applies as the data is aggregated nationally. This data source is useful for estimating how much the Medicare Part B Activity file omits due to the suppression rule.

Medicare Part D Spending by Drug – This dataset is the companion file to the Medicare Part D Activity file, reporting activity by branded-generic name but aggregated at the national level. The "10 or fewer beneficiaries" suppression rule still applies but rarely takes effect due to the national aggregation. This data source is valuable for assessing how much the Medicare Part D Activity file excludes because of the suppression rule.

C. Insurance Status

U.S. Census ACS (American Community Survey) – This is the go-to data source when it comes to health insurance status. This is because the data is collected at the individual household member level and indicates whether

or not the person has insurance. When a person has insurance, the dataset specifies whether the insurance is private (employer-based or direct-purchase) or public (Medicare, Medicaid, Tricare, VA, Indian Health Service, etc.).

Medicare Enrollment – CMS (Centers for Medicare & Medicaid Services) maintains two separate data sources: one for Original Medicare and another for Medicare Advantage. Both databases provide detailed enrollment information, including the number of enrollees for each county, broken down by plan and contract.

SHADAC (State Health Access Data Assistance Center) – This is another data source worth mentioning, although it lacks the requisite county-level granularity and only reports health insurance status data at the state level. The data is presented in a straightforward and user-friendly manner. Note that SHADAC does not produce the data itself but aggregates existing federal surveys such as the ACS (American Community Survey), CPS (Current Population Survey), MEPS (Medical Expenditure Panel Survey), and NHIS (National Health Interview Survey).

State Health Facts Tool from KFF (Kaiser Family Foundation) – This data source offers insurance coverage data at the state level for both adults and children, differentiating between public and private coverage. It also covers aspects of access and affordability—such as out-of-pocket expenses, cost-related barriers to care, and metrics related to medical debt and dental visits. Like SHADAC, KFF aggregates its data from federal surveys such as the ACS, CPS, MEPS, and NHIS. In addition to these primary data sources, KFF also incorporates administrative data from Medicare and Medicaid.

Commonwealth Fund – This is yet another data source that reports data at the state level but along different, though related, dimensions: (1) Access and Affordability (insurance coverage, costs, and barriers), (2) Prevention and Treatment (use of preventive services and care quality), (3) Potentially Avoidable Hospital Use and Cost (unnecessary hospital and emergency usage with associated spending), (4) Healthy Lives (overall health outcomes and risk behaviors), and (5) Reproductive Care and Women's Health (maternal, infant, and women's service outcomes, sometimes detailed by race and ethnicity).

Use Case 4 - Development of Look-Alike Models for Expanding Therapy Usage

Objective: Expand drug adoption and usage.

Approach: Focus not on the providers who use our drugs, but rather on the environment in which these providers operate. Zero in on environments similar to those where we have experienced great success with providers, and avoid environments that resemble those where we have not. In short, focus not on the fish, but on the fish tank.

Geographic Granularity – The county arguably has the right level of resolution to capture the environment. This is for two reasons. First, the zip code is too granular. Indeed, in countless instances, hospitals, medical practices, and other providers may operate in one zip code while the patients they serve live in nearby residential zip codes. This means that operating at the zip level would lead us to compare professional and residential zip codes as if they were separate when they belong to the same community. Second, many high-quality data sources report data at the county level. Each state has, on average, 60+ counties, and a county has about 106,000 people on average.

These statistics are based on 2020 data: 3,144 counties and 333.29 million people.

From a medical standpoint, there are 4 dynamics at play.

1. Demand – The number of people who have the condition or disease that may benefit from our therapy.

2. Supply – Physicians and hospitals that can deliver the therapy we recommend.

3. Lubricant – Payers are the predominant lubricant between patients and providers and play a crucial role in ensuring that supply aligns with demand. Without their approval, patients cannot receive the therapy prescribed by providers.

4. Backdrop – This refers to the fabric as well as the subtle forces at play in the community. These include the types of jobs people have, level of education, income and wealth, exercise and drinking habits, religious beliefs, access to healthcare, walkability, presence of parks, level of pollution, amount of sunshine (as in levels of UV-A and UV-B), and similar factors.

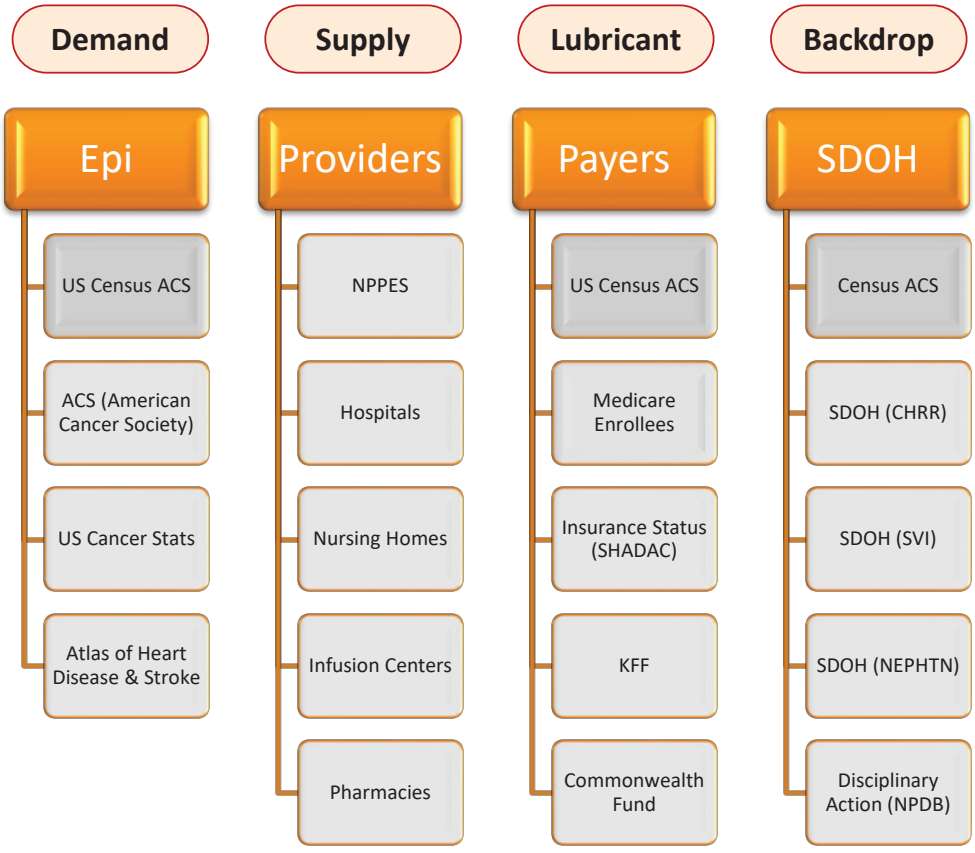


Fig 5 – List of Data Sources in the Public Domain for Developing Look-Alike Models

A. Demand Data Sources – Epidemiology

The go-to data here is county-level epidemiology data (incidence, prevalence, and mortality). At the top of the list is the US Census ACS (American Community Survey), as it offers detailed data on population size, age, racial composition, income, and education levels, among other variables. It reports data at the county level and also at the ZIP level and even lower, at the census tract level, depending on the variable.

If the disease area we are looking at is cancer, ACS (American Cancer Society) and US Cancer Stats are excellent choices. SEER, although a great source of reliable data, is not our first choice, as it reports data for only 37% of the US population.

If the disease area is cardiovascular, as in coronary heart disease, stroke, heart failure, hypertensive heart disease, arrhythmias, peripheral artery disease, congenital heart disease, valvular heart disease, rheumatic heart disease, cardiomyopathies, and the like, we recommend the Atlas of Heart Disease and Stroke, which is the result of a collaboration between WHO (World Health Organization) and CDC (Centers for Disease Control and Prevention).

B. Supply Data Sources – Physicians, Hospitals, Care Centers, and Pharmacies

NPPES (National Plan and Provider Enumeration System) – This data source lists all 7 million providers: 5 million individual providers, including physicians, NP/PAs, nurse practitioners, therapists, and other licensed practitioners, and 2 million institutional providers, including hospitals, skilled nursing facilities, ambulatory care facilities, pharmacies, home health agencies and hospices, labs, and the like. NPPES indicates, for each provider, the name, address, city, state, and ZIP, among other things, which allows us to get a count of the number of providers by specialty for each county. It is managed by the Centers for Medicare & Medicaid Services (CMS), a division of the U.S. Department of Health and Human Services, and is updated on a weekly basis.

AHA (American Hospital Association) Fast Facts and AHD (American Hospital Directory) – These data sources allow us to paint a more nuanced picture of providers as they provide information on the number of beds, staffing levels, number of admissions, outpatient visits, lengths of stay, revenues and expenses, EHR adoption, and the like.

Medicare Care Compare – We may also leverage data sources from the Medicare Care Compare platform, where we can compare not only hospitals but also nursing homes (skilled nursing facilities), home health agencies, hospices, inpatient rehabilitation facilities (IRFs), and long-term care hospitals (LTCHs). Each type of provider has its own set of quality metrics and inspection results. Some of the variables include an overall star rating, clinical outcomes, inspection results, staffing levels, and patient experience.

NICD (National Infusion Center Directory) – This is a data source that is maintained by NICA (National Infusion Center Association) based on self-reported information from pharmaceutical companies. It lists all the outpatient infusion centers along with their locations and the infused drugs that are administered at the site, among other things. This can help capture the outpatient infusion landscape at the county level.

Pharmacy Network – This is a data source that tracks pharmacies participating in Medicare's prescription drug programs (e.g., Medicare Part D) and allows us to go beyond

NPPES by shedding light on the drugs the pharmacy carries. Pharmacy Network is a companion data source to FRF (Formulary Reference File) and SCC (State County Contracts) from CMS and describes the 800+ contracts between Medicare and individual pharmacies (referenced by an NPI) at the county level. Indeed, it spells out the terms of the formulary in place for each drug, which can then be used to zero in on the relevant pharmacies when portraying the pharmacy picture at the county level for a given disease or therapeutic area.

C. Lubricant Data Sources – Health Insurance Status

US Census ACS (American Community Survey) – This is the go-to data source when it comes to health insurance status. That's because the data is collected at the individual household member level and indicates whether the person has insurance. When the person has insurance, it further specifies whether the insurance is private (employer-based or direct-purchase) or public (Medicare, Medicaid, Tricare, VA, Indian Health Service, etc.).

Medicare Enrollment – CMS (Centers for Medicare & Medicaid Services) has two separate data sources: one for Original Medicare and another for Medicare Advantage. Both databases provide detailed information on enrollment, including the number of enrollees for each county, broken down by plan and contract.

SHADAC (State Health Access Data Assistance Center) – This is another data source worth mentioning, although it lacks the requisite county granularity and only reports health insurance status data at the state level. The data is presented in a very straightforward and easy-to-use manner. Note that SHADAC does not produce the data but rather aggregates existing federal surveys such as ACS (American Community Survey), CPS (Current Population Survey), MEPS (Medical Expenditure Panel Survey), and NHIS (National Health Interview Survey).

State Health Facts tool from KFF (Kaiser Family Foundation) – This is another data source that offers insurance coverage data at the state level for both adults and children, differentiating between public and private coverage. It also covers aspects of access and affordability — such as out-of-pocket expenses, cost-related barriers to care, and metrics related to medical debt and dental visits. Like SHADAC, KFF aggregates its data from federal surveys such as ACS (American Community Survey), CPS (Current Population Survey), MEPS (Medical Expenditure Panel Survey), and NHIS (National Health Interview Survey). In addition to these primary data sources, KFF also incorporates administrative data from Medicare and Medicaid.

Commonwealth Fund – This is yet another data source that also reports data at the state level but along different, though related, dimensions: (1) Access and Affordability (insurance coverage, costs, and barriers); (2) Prevention and Treatment (use of preventive services and care quality); (3) Potentially Avoidable Hospital Use and Cost (unnecessary hospital and emergency usage with associated spending); (4) Healthy Lives (overall health outcomes and risk behaviors); and (5) Reproductive Care and Women's Health (maternal, infant, and women's service outcomes, sometimes detailed by race and ethnicity).

D. Backdrop Data Sources – SDOH (Social Determinants of Health)

CHRR (County Health Rankings & Roadmaps) – This is a very rich and powerful

SDOH data source at the county level. It is a collaborative effort led by the Robert Wood Johnson Foundation (RWJF) in partnership with the University of Wisconsin Population Health Institute (UWPHI) and incorporates data from sources such as BRFSS (Behavioral Risk Factor Surveillance System) to assess and compare the overall health of U.S. counties. Under its health factors, it includes variables spanning the physical environment, social and economic factors, clinical care, and health behaviors. In its core rankings, CHRR measures indicators such as smoking rates, preventable hospital stays, income inequality, and air quality, which collectively inform overall county health outcomes, such as length of life and quality of life.

PLACES (Population Level Analysis and Community Estimates) – This CDC resource is also a valuable data source that reports data at the county level. It leverages BRFSS (Behavioral Risk Factor Surveillance System) data in collaboration with UWPHI (University of Wisconsin Population Health Institute) and focuses on generating modeled, small-area estimates of health outcomes, risk factors, and access to care at the county (and even sub-county) level. It reports data under four areas: (1) Disability (cognitive, hearing, vision, etc.), (2) Prevention (annual checkup, dental visit, mammography, etc.), (3) Health Outcomes (obesity, high blood pressure, depression, etc.), and (4) Health Risk Factors (smoking, binge drinking, physical inactivity, etc.).

SVI (Social Vulnerability Index) – This is a county-level data source developed by the CDC (Centers for Disease Control and Prevention), specifically through its ATSDR (Agency for Toxic Substances and Disease Registry). It uses U.S. Census data to identify communities that may be more vulnerable to external stresses such as natural disasters, disease outbreaks, or other emergencies. It assesses vulnerability along four dimensions: (1) Socioeconomic Status (unemployed, no high school diploma, no health insurance, etc.), (2) Household Characteristics (civilian with disability, single-parent households, English language proficiency, etc.), (3) Racial & Ethnic Minority Status (Hispanic, Black, Asian, etc.), and (4) Housing Type & Transportation (mobile homes, no vehicle, crowding, etc.).

NEPHTN (National Environmental Public Health Tracking Network) – This is a very valuable data source as it connects health (or the lack thereof) with environmental factors. It is managed by the CDC (Centers for Disease Control and Prevention) through a collaborative effort between the CDC's NCEH (National Center for Environmental Health) and ATSDR (Agency for Toxic Substances and Disease Registry). This initiative integrates data from state, local, tribal, and territorial partners to monitor environmental hazards and related health outcomes. It tracks variables such as air quality, childhood lead poisoning, drinking water, drought, pesticide exposure, precipitation & flooding, radon, sunlight & UV, tornadoes, release of toxic substances, unintended carbon monoxide poisoning, and more.

NPDB (National Practitioner Data Bank) – This is a federal database of reports on adverse actions like malpractice payments, licensure issues, and clinical privileges. It is maintained by the U.S. Department of Health and Human Services Office of Inspector General. Full disclosure of the identity of physicians is available to hospitals and physicians but not to the general public (the NPI is replaced by an internal physician ID). This data source offers information at the state level and serves as yet another valuable yardstick for comparing geographies.

Key Takeaways

Here are the key takeaways we'd like to leave you with.

1. Expand your list of go-to data sources. Include free data sources, as many of them are of high quality. They serve two purposes. First, they help answer business questions that we would otherwise not be able to address. Open Payments from the Sunshine Act, for example, is the only data source that sheds light on the amount of money physicians receive from pharmaceutical companies. Second, they help assess the quality of other data sources. They serve as yardsticks and help us identify holes, biases, and uncover caveats we would otherwise miss.

2. You're not alone in being unaware of these data sources; their obscurity stems precisely from the fact that they are free. We call this the "Paradox of the Free." Because these data sources are free, there's no one to promote them or answer questions about them. As a result, they remain dormant until someone stumbles upon them and spreads the word.

3. Think outside the box and look beyond advertised uses. For instance, Open Payments from the Sunshine Act allows us to get a feel for the activity of reps who call on physicians to promote specific drugs across the U.S. Here's another example. Transparency in Coverage (TIC) allows us to gauge the business acumen of physicians, as it reports all the payer networks the physicians belong to and the corresponding negotiated reimbursement amounts for each procedure.

As of this writing, the workforce of federal agencies is being gutted one after another. It is very possible that some of the data sources we mentioned may no longer be updated, or worse, may be taken down entirely. We can only hope this does not happen.

Resources

1. CMS data: https://data.cms.gov/
2. Open Payments: https://openpaymentsdata.cms.gov/
3. National Environmental Public Health Tracking Network (EPH) https://ephtracking.cdc.gov/
4. County Health Rankings & Roadmaps (CHRR): https://www.countyhealthrankings.org/
5. CDC Fast Facts from National Center for Health Statistics: https://www.cdc.gov/nchs/fastats/default.htm
6. American Community Survey (ACS) data from Census: https://www.census.gov/programs-surveys/acs
7. Social Vulnerability Index (SVI) from Centers for Disease Control (CDC) and Prevention and Agency for Toxic Substances and Disease Registry (ATSDR): https://www.atsdr.cdc.gov/place-health/php/svi/index.html
8. Surveillance, Epidemiology, and End Results (SEER) from NIH National Cancer Institute: https://seer.cancer.gov/
9. US Cancer Statistics from CDC: https://www.cdc.gov/united-states-cancer-statistics/
10. National Practitioner Data Bank (NPDB) from Health Resources and Services Administration (HRSA): https://www.npdb.hrsa.gov/index.jsp
11. MedPAR LDS: https://www.cms.gov/data-research/files-for-order/limited-data-set-lds-files/medpar-limited-data-set-lds-hospital-national

About the Authors

Jean-Patrick Tsang is the Founder and President of Bayser, a Chicago-based consulting firm dedicated to sales and marketing for pharmaceutical companies. JP is an expert in data strategy and advanced analytics. JP has published 25+ papers, given 80+ talks at conferences, and completed 250+ projects. In a previous life, JP deployed Artificial Intelligence to automate the design of payloads for satellites. JP earned a Ph.D. in Artificial Intelligence from Grenoble University, advised 2 PhD students, and earned an MBA from INSEAD in Fontainebleau, France. He was the recipient of the 2015 PMSA Lifetime Achievement Award.

Shunmugam Mohan is a Vice President with Bayser. His expertise has been in developing Portfolio/Data Strategy for Oncology and Rare Disease Markets using a myriad of data sources ranging from open/closed PLD, SP, retail pharmacy, lab, NGS, EMR, health exchange, remit, hierarchy/affiliation, CMS and SDOH. He finds passion in looking out for new data sources to effectively answer business questions based on their strengths and caveats. Prior to joining Bayser, Shunmugam worked in immune system modeling where he specialized in 3D texture analysis of multiple imaging modalities during his Master's in Electrical Engineering at the Ohio State University. He can be reached at (224) 365-9462 or shunmugam@bayser.com.

Zhangjin Xu is a data science professional with expertise in healthcare analytics, clinical trial optimization, and pharmaceutical data strategy at Bayser. Her work focuses on leveraging machine learning, social determinants of health, and generative AI tools like ChatGPT to enhance site selection, data quality, and predictive modeling. Zhangjin has developed scalable solutions for data merging, anomaly detection, and semantic data assessment, enabling more informed and efficient decision-making in clinical and commercial settings. Prior to joining Bayser, Zhangjin worked in theoretical physics where she specialized in nonlinear optics modalities during her PhD's at Mississippi State University.

Maylis Larroque is a senior consultant with Bayser, and her expertise lies in identifying and processing data sources across different therapeutic areas and countries to help clients answer business questions. She spearheaded the development of Panorama, a large database of health social determinants of health with over 1,000 variables. Previously, Maylis worked in GMP environments, quality assurance investigations and bio-contamination control. Maylis holds a Master's Degree in Pharmaceutical Sciences from Bordeaux University in France. She can be reached at maylis@bayser.com.

Looking Beyond Hype: Using AI to Drive Business Impact for Brand and Sales Leaders

Vineet Rathi, Principal, Axtria Inc.

Abstract: Artificial intelligence (AI) has been part of the pharmaceutical industry for several years. However, some life sciences companies remain stuck and have yet to find a way to move beyond the hype. Now, they find themselves with AI initiatives seemingly destined to fail. This has become an all too familiar and critical issue, thanks to an increasingly competitive healthcare landscape. Moving beyond the hype requires a strategic approach to AI adoption that ensures sustainable success.

More specifically, brand and sales leaders are under immense pressure to be more productive and to make better-informed decisions. They know AI is a game-changer, but without proof of a successful program, it becomes difficult to justify its use. In this paper, the author examines how pharmaceutical leaders can prove success by implementing AI correctly from the get-go, thereby leveraging it to drive meaningful business impact and finally moving beyond the hype.

Introduction

As brand and sales leaders face increasing pressure to enhance productivity, improve decision-making, and deliver measurable outcomes, artificial intelligence (AI) has emerged as a potential game-changer. However, the jury is still out on the effectiveness and business impact that AI is driving. Navigating the landscape of AI adoption often entails separating hype from actionable opportunities. This paper examines how pharmaceutical leaders can leverage AI to drive meaningful business impact, moving beyond exaggerated claims to implement strategic, data-driven solutions.

Understanding AI's Role in Pharmaceuticals

Within the pharmaceutical sector, AI applications now span a myriad of use cases, including drug discovery, clinical trials, marketing, sales optimization, and patient engagement. No matter what the use case is, business leaders must show a reason why they need AI. To do that, they must identify key drivers, relay the significance, recognize the differences in use cases, and understand the barriers to implementation and adoption.

Key Drivers

There are several forces of change in the pharmaceutical industry: the emergence of more specialty therapies, the increased focus on commercial design strategies for organized customer groups, changing market access dynamics, enhanced leverage of patient-focused data and insights, and pressures due to the Inflation Reduction Act, among others. These drivers mean leaders must be ready to solve three critical needs in any life sciences undertaking – speed, personalization, and intelligence.

> ***Need for "Speed"*** is critical throughout the product lifecycle, from research and development to launch and throughout the exclusivity period. Increased pressure

to reduce the timelines in each phase of the product lifecycle results in companies turning their focus on AI as a key lever to accelerate the entire process.

Speed is also crucial to meeting customer and stakeholder needs throughout the value chain. Patients, providers, and payers are all demanding agility and higher responsiveness throughout the continuum of the patient journey and all related healthcare processes, making AI a critical need.

Need for "Personalization" of the customer and patient experience has become a must-have. Without exception, the field force must know what customers and patients need, when they need it, and how to deliver it to a "segment of one." The variety of customer types (e.g., HCPs, hospitals, IDNs, payers) further amplifies the need to bring personalization in customer engagement strategies given the unique objectives and priorities of different customer stakeholders. The same is true of treatment plans. Gone are the days of one-size-fits-all therapies. The technological advancements made possible by AI have enabled the development of personalized therapies and treatment plans designed for a patient's specific needs.

Need for "Intelligence" is becoming critical as there is a vast amount of data getting generated in the healthcare ecosystem, and stakeholders are getting overwhelmed with the flood of information, which often providing conflicting or incomplete information. The challenge lies in finding ways to extract actionable intelligence from it and establish "trust" that its insights are valid. The need for agile intelligence when making decisions is becoming critical for all pharmaceutical stakeholders, and that is amplifying the importance of AI.These key drivers have created a more urgent need for companies to leverage AI across the entire value chain, from pre-discovery to post-discovery. Life sciences leaders have seen the evidence that, compared to traditional methods, AI-infused approaches can bring significant agility and added value.

Significance

Pharma sales and brand leaders are looking beyond the hype. They are demanding AI use cases they can industrialize to drive measurable impact and business outcomes. The reason is clear – Pharma companies that industrialize AI use cases across their organizations have the potential to double their operating profit.

The expected AI/ML - and GenAI-driven gain in the operating profits of pharma companies worldwide by 2030 is over $250 Billion.[1]

Key Applications of AI for Brand and Sales Leaders

For brand and sales leaders, AI can be a powerful tool for personalizing marketing strategies, optimizing sales operations, and enhancing patient and customer engagement. Key applications for AI in the pharma commercial space include:

- **Predictive Analytics**
  Predictive analytics leverages historical data and various influencing parameters to forecast future trends and behaviors. By analyzing physician prescribing patterns, patient demographics, and market dynamics, AI tools can predict which products are likely to succeed in specific regions. The results enable sales teams to allocate resources more efficiently

and prioritize high-potential accounts. Moreover, real-time updates help adjust strategies dynamically, ensuring optimal outcomes.

Selecting the right AI model for various predictive analytics tasks is key to managing the temporal nature of pharmaceutical data, which often involves complex, time-dependent patterns. Pharmaceutical sales data, for example, may be influenced by factors such as seasonal trends, regulatory changes, or product life cycle stages, necessitating models that can effectively capture these dynamics. When dealing with rare diseases or specialty drug launches, where data is sparse or events are infrequent, strategies like transfer learning, anomaly detection, or synthetic data generation can help enhance model performance. Additionally, incorporating market events and competitive dynamics into AI models requires integrating real-time external data, such as competitor activities, market shifts, or healthcare policy changes. Techniques like causal inference and scenario modeling can help assess the impact of these factors on sales forecasts, allowing pharmaceutical companies to adapt to changing market conditions and optimize their commercial strategies.

- **Patient and Customer Segmentation**
  AI-driven segmentation surpasses traditional demographic-based approaches by incorporating behavioral, psychographic, and contextual data. Advanced clustering algorithms can uncover hidden patient and customer segments, enabling tailored targeting and messaging that resonates with specific needs. Generative AI (GenAI) can further personalize outreach by creating custom messaging templates for distinct patient and customer segments, ensuring alignment with their unique preferences.

- **Productivity Enhancement for Field Roles**
  AI- and GenAI-powered insights and coaching guidance are particularly useful levers for improving the effectiveness of one of the most significant investments most pharmaceutical companies make—their field roles. AI-powered analytics can provide real-time insights into healthcare provider preferences, prescribing behaviors, and patient outcomes, enabling sales reps to personalize their interactions and tailor their pitches to specific needs. Additionally, AI tools can optimize territory management, call planning and dynamic targeting helping reps prioritize high-potential accounts and allocate resources more efficiently. As a result, AI not only increases the efficiency and effectiveness of field roles but also enhances the overall impact of sales and medical outreach, driving better results for pharmaceutical companies.

- **Sentiment Analysis**
  Understanding customer sentiment is crucial for pharmaceutical brands aiming to build trust and loyalty. Sentiment analysis uses natural language processing to interpret text data from surveys, reviews, or online platforms. By identifying positive, neutral, or negative sentiments, brands can proactively gauge public perception and address concerns. GenAI tools enhance sentiment analysis by generating detailed summaries of trends and proposing responses to negative feedback, improving crisis management strategies.

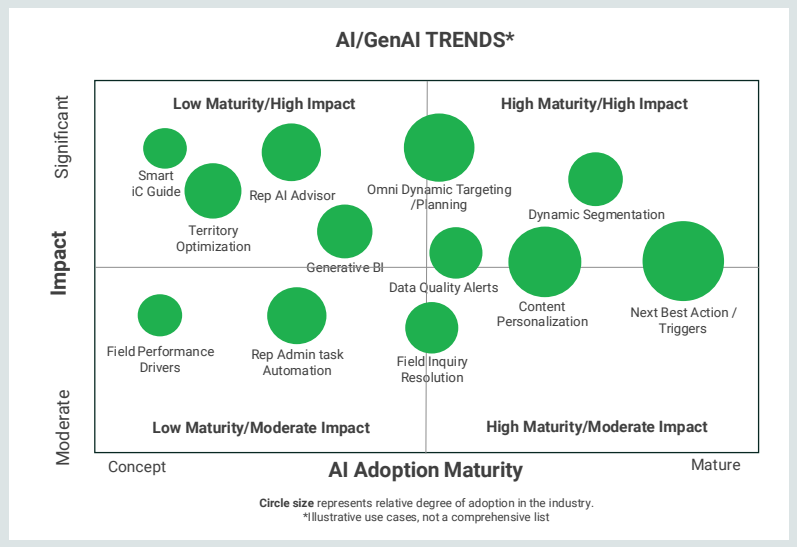
- **Chatbots and Virtual Coaches**
  AI-powered chatbots are evolving from supporting basic inquiries to more

advanced virtual coaches, able to help different user personas, such as field representatives, sales and brand leaders, or commercial strategy and operations teams. These virtual coaches help users understand the "why" behind the insights and provide directional guidance on potential next actions. These tools enhance customer experience and free up human resources for high-value tasks. GenAI technologies have advanced chatbot capabilities, enabling more natural and context-aware conversations.

- **Competitive Intelligence**
  AI tools aggregate and analyze market data to provide competitive intelligence, helping brands stay ahead in a dynamic environment. Pharmaceutical leaders can adapt their tactics by monitoring competitor activities, pricing strategies, and promotional campaigns to maintain a competitive edge. Generative AI can assist in generating detailed competitor analysis reports, reducing the time required for manual research.

The list above is by no means exhaustive, and many new use cases emerge daily; however, pharma and life sciences leaders must remember – not all use cases are equal. This illustration shows how some use cases in the industry are faring with user adoption:

AI Trends in Pharma Commercial Excellence Value Chain



Key Barriers to Implementation and Adoption

Despite its transformative potential, the industry must know how to address challenges when integrating AI into operations. These roadblocks run the gamut from data silos to regulatory compliance to ethical considerations. In fact, some AI initiatives at pharma companies fail to move beyond the proof of concept or pilot stage for several reasons:

1. **Lack of a well-defined business objectives and lack of stakeholder alignment**

   AI programs and use cases are often created without full due diligence around expected business outcomes. Putting more focus on technology leads to complexity, which makes it cumbersome for key business stakeholders, like field, sales, and brand leadership, to understand the objectives and outcomes. The industry needs to avoid taking shortcuts due to the "fear of missing out" phenomenon.

2. **Data quality and integration**

   AI systems rely on high-quality, structured data for accurate predictions. However, pharmaceutical companies often face data silos and inconsistencies across departments. Integrating diverse sales, marketing, and clinical operations datasets into a unified platform requires significant investment and expertise.

   The use of AI in pharmaceutical commercial analytics faces several data-related challenges that need to be addressed for effective implementation. One key requirement is the establishment of robust data architectures, such as data lakes, that can handle the integration of real-time data from diverse sources like sales transactions, customer interactions, market access and external market data. Pharmaceutical companies need to manage large volumes of structured and unstructured data, ensuring that it is easily accessible and standardized for AI model training and analysis. A data lake architecture allows for the storage of both raw and processed data, facilitating real-time analytics and decision-making. However, integrating real-time data from multiple sources often involves overcoming challenges related to data silos, inconsistent formats, and varying update frequencies. Another challenge is managing unstructured data, especially from field interactions like sales reps' notes, emails, or voice recordings. Natural language processing (NLP) and advanced text analytics can be applied to extract valuable insights from this unstructured data, but ensuring data quality and relevance remains a significant hurdle. Strategies like data labeling, categorization, and the use of AI to detect patterns in field interactions can help organizations unlock the value of this unstructured data and incorporate it into commercial decision-making.

   In today's world of Generative AI, a company's data strategy must also focus on building and maintaining GenAI-specific data assets and structures. To build GenAI-Ready Datasets (GRDs), one must also design and develop the right processes around creating, maintaining, and tweaking/enriching such GRDs over time.

   Addressing these data challenges is critical for the successful use of AI in pharmaceutical commercial analytics, enabling more accurate predictions and enhanced business strategies.

3. **Regulatory compliance**

   Pharmaceutical companies operate under stringent regulatory frameworks. The use of AI must comply with ironclad governmental rules, such as the Health

Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union. Further regulations and guidelines by the U.S. Food and Drug Administration (FDA) ensure data privacy and ethical practices. Non-compliance can lead to high-cost legal implications and an overall risk to brand perception.

Addressing the challenges of using AI in the pharmaceutical industry requires careful attention to data governance, particularly when it comes to maintaining data lineage for regulatory compliance. Companies must ensure that their data is traceable, auditable, and accurate to meet strict compliance standards set by regulatory bodies. Data lineage, which tracks the origin, movement, and transformations of data throughout its lifecycle, is crucial to demonstrate transparency and accountability in AI-driven decisions, especially when those decisions impact drug development, sales, or patient safety. Failing to maintain proper data lineage can lead to compliance risks, legal challenges, and damage to the company's reputation. As a result, investing in the right data platform is essential. A robust platform that supports data lineage not only enables seamless tracking of data but also ensures data integrity, facilitates audits, and simplifies compliance reporting. Additionally, such platforms allow for better integration and management of diverse data sources, ensuring that pharmaceutical companies can leverage AI while meeting the necessary regulatory requirements.

4. Change management

Introducing AI into pharmaceutical sales and brand management requires a cultural transformation within the organization. Resistance to change is common, particularly due to concerns over accuracy, a lack of familiarity with emerging technologies or job displacement. To overcome this, effective change management strategies are crucial. This includes comprehensive training programs to upskill employees, along with proactive communication and stakeholder engagement to build trust and alignment. A key component of this trust-building is ensuring transparency and clarity around how AI-generated insights and recommendations are derived. Demonstrating the reliability and accuracy of AI-driven results through clear validation processes can foster confidence among sales and brand stakeholders. By addressing concerns, educating on AI's role, and showcasing its value, organizations can drive adoption, ensuring that AI integration enhances decision-making and maximizes return on investment, rather than creating fear or disruption.

5. Algorithm transparency

AI algorithms often function as "black boxes," making it challenging to interpret their decision-making processes. This lack of transparency raises concerns among stakeholders, particularly in a highly regulated industry like pharmaceuticals. Ensuring algorithm explainability is crucial to building trust and enabling informed decision-making, particularly when these models are used to make critical commercial recommendations.

Implementing robust model governance frameworks is necessary to strike a balance between technical accuracy and regulatory compliance. These frameworks typically include clear guidelines on model development, validation, and deployment,

ensuring that AI models not only meet the required performance standards but also adhere to industry regulations. Additionally, companies may employ validation protocols like cross-validation, sensitivity analysis, and explainability methods to assess model reliability and interpretability. Standard metrics, such as errors, variances, mean absolute error, etc., should be tracked, and models should be tweaked as needed. This helps ensure that commercial recommendations are based on sound data and are justifiable in the context of regulatory scrutiny. Validation processes might also include testing models against historical data, comparing predictions with real-world outcomes, and conducting SME (subject matter expert) reviews, all of which contribute to building trust in the AI system. By adhering to these governance and validation protocols, pharmaceutical companies can foster confidence in their AI-driven decisions and mitigate potential risks related to transparency and compliance.

Overall, lessons from the AI journeys of multiple pharmaceutical companies can help reduce implementation and adoption barriers. Lessons learned include:

- **Privacy matters:** Ensuring data privacy is a top priority, especially when dealing with sensitive health and patient information. Maintaining compliance with regulations like GDPR and HIPAA is crucial to avoiding legal issues and building trust with stakeholders.

- **Misaligned incentives:** There can be a disconnect between various stakeholders (e.g., data scientists, business leaders, and regulatory bodies), leading to conflicting goals. Aligning incentives early in the process ensures smoother project execution and a shared understanding of objectives.

- **Preference for simplicity and common sense:** While AI can be powerful, simple solutions often yield the best results in pharmaceutical commercial projects. Complex algorithms or overly technical approaches can backfire if they aren't intuitive or they fail to address core business needs.

- **Shiny New Object Syndrome:** It's easy to get distracted by the latest AI trends or technologies. However, focusing on real-world applicability and proven tools rather than chasing the "next big thing" is vital for achieving practical and sustainable outcomes.

- **Rushing into an implementation without ensuring adequate infrastructure capabilities and governance exist:** Implementing AI without a solid technological infrastructure or governance framework can lead to inefficiencies, errors, and security risks. Ensuring robust systems are in place first will pave the way for smoother, more effective deployment.

- **Focus on customer-centered engagement:** AI should enhance, not replace, customer interactions. Keeping the customer experience at the forefront ensures that AI tools add real value by improving engagement, personalization, and service delivery.

Strategies for Successful AI Adoption

Clearing the hurdles outlined above is still only half the AI battle. Companies cannot justify an AI initiative unless it can be industrialized and scaled across adjacent departments, brands, and use cases. Only then can the true business impact of AI be realized. To reach that stage,

firms must take advantage of a blueprint for success – a concise, step-by-step strategy for scalability and adoption.

1. Define clear business objectives

Pharmaceutical leaders must identify specific business problems that AI can address. Whether improving sales forecasting or enhancing customer engagement, clearly defined objectives ensure focused efforts and measurable outcomes.

2. Foster collaboration

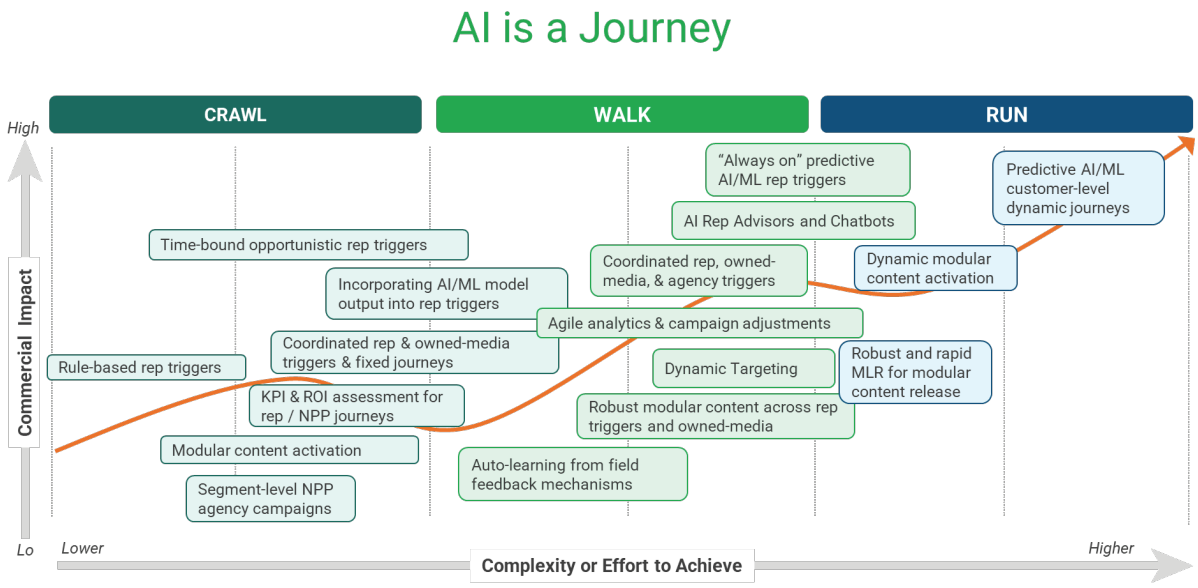
Effective cross-functional collaboration between sales, brand, marketing, IT, and market access teams is vital to aligning AI initiatives with broader business objectives. Creating interdisciplinary task forces helps break down silos, ensuring seamless communication and cooperation. This approach not only streamlines the AI implementation process but also drives scalability and accelerates adoption across the organization, ensuring that AI solutions are integrated effectively and deliver maximum value.

3. Choose the right AI partner

Collaborating with reputable AI vendors or consulting firms can accelerate adoption while minimizing risks. Evaluating potential partners based on their consulting and platform expertise, track record, and compliance standards is critical to ensuring long-term success.

4. Create a business outcome focused roadmap

Creating a well-defined roadmap will ensure that each AI initiative is tied to the overall big picture and company objectives. Having a roadmap also helps define measurable picture of success and ensure that corrective actions are taken sooner rather than later. An example of a typical AI journey roadmap for targeting is shown below:



Source: Axtria Inc.

Operationalizing AI at Scale: Implementation Framework

For AI to be successfully implemented, a structured framework is essential to ensure the accuracy, compliance, and scalability of AI models. This framework should cover key areas such as model development, data infrastructure, system integration, testing, and performance monitoring.

1. Model Development Lifecycle Management

A structured approach to the model development lifecycle (MDLC) is critical to ensure AI models are developed, tested, deployed, and maintained effectively.

- Model Design & Selection: Choose appropriate AI techniques (e.g., deep learning, reinforcement learning) depending on the task (e.g., classification, regression, prediction).
- Model Training & Optimization: Train the model using large-scale datasets and optimize it using techniques like hyperparameter tuning, cross-validation, etc. Once a model is in production and is being used, over time, to sustain or better the quality of outputs, it needs to be (re) trained, fine-tuned (possibly needing adjustments to hyperparameters), etc.
- Model Evaluation: Evaluate model performance using key metrics (e.g., precision, recall, F1-score) and validate it using unseen data to ensure robustness and generalization.
- Computationally intensive models need specialized hardware (GPUs, TPUs, etc.) to bring response times to an acceptable level, while others work off regular conventional hardware. Of course, this leads to cost, resourcing, and other factors. These are important factors that must be considered when selecting a model.

2. Data infrastructure

Building a robust data infrastructure is foundational to AI success. A key component of this is data acquisition, which plays a critical role in shaping the data strategy. Ensuring access to the right data—whether from internal systems, external sources, or real-time inputs—is vital for AI success. Alongside data acquisition, data cleansing, integration, lineage and governance practices are essential to maintain high-quality, actionable data.

Pharmaceutical companies must focus on not just acquiring data, but also on extracting optimal value from their data investments. By implementing strong data governance and integration frameworks, organizations can ensure that their AI initiatives are fueled by accurate, timely, and relevant data, maximizing their potential impact, adoption and scalability

3. Intuitive user interface

Designing an intuitive UI for AI models requires tailoring the interface to the needs of different user types, such as power users and regular business users, by leveraging principles of user-centered design and information architecture. For power users, who are more familiar with data and model intricacies, the UI should support advanced features like customizable dashboards, data drill-downs, and interactive visualizations that allow users to explore model predictions, adjust parameters, and perform complex queries through dynamic interfaces. It should

incorporate data widgets, filtering options, and advanced analytics tools for in-depth interaction. In contrast, for regular business users, the interface should focus on simplicity and clarity, offering data-driven visualizations with intuitive navigation. Use of call-to-action buttons, tooltips, and onboarding tutorials can help guide users without overwhelming them with technical jargon. Employing a responsive design ensures that the UI adapts across different devices. Additionally, applying iterative design through user feedback loops and conducting regular usability testing ensures the interface evolves based on user needs, enhancing accessibility and adoption for all user levels.

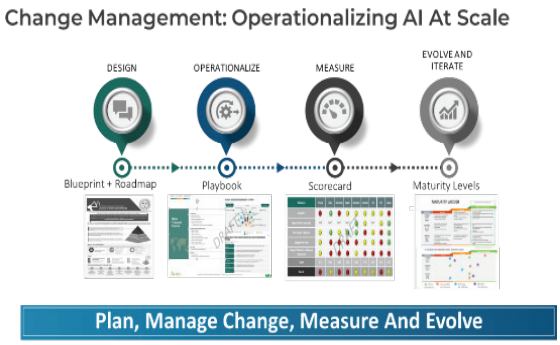
4.System Integration

Integrating AI solutions with existing systems is crucial for seamless operation and minimal disruption. Proper integration allows AI models to ingest and process data, perform predictions, and send results back to the relevant systems, making AI truly functional within the organization. System Integration techniques include API-based integration, batch integration, cloud service integration and microservices architecture i.e. develop AI models as microservices, making it easier to scale and integrate with existing systems. This approach allows for modular development, testing, and deployment.

5.Testing and Performance Monitoring

AI model testing best practices involve a systematic approach to ensure accuracy, reliability, and compliance. First, it's essential to define clear testing objectives, including performance metrics like precision, recall, and F1 score, based on the model's intended use. Cross-validation should be employed to evaluate generalizability and prevent overfitting. Testing should include edge cases and real-world scenarios to assess robustness. It's crucial to test for bias and fairness, ensuring the model doesn't produce discriminatory or unethical results. Additionally, thorough integration testing ensures the model works seamlessly with existing systems and data pipelines.

Finally, continuous monitoring, performance tracking and change management should be established post-deployment to detect model drift and allow for periodic updates and retraining. Regular performance evaluations and updates ensure that the model remains aligned with business objectives and regulatory requirements, fostering long-term reliability and trust.



Source: Axtria Inc.

Business Outcomes and Conclusion

Companies that have been able to leverage AI best practices have seen significant business benefits across several dimensions. The image illustrates some of the benefits Axtria has seen across its partner pharmaceutical companies and highlights the potentially significant value that can be unlocked.

Source: Axtria Inc.

Quantitative success metrics are essential for measuring AI implementation success, offering concrete insights into both performance and business impact. Model performance metrics such as precision, recall, F1 score, and accuracy assess how well the AI model predicts outcomes and avoids errors. System performance indicators like latency, throughput, and error rates measure the efficiency and reliability of the model in production. User adoption metrics, including active users, session duration, and retention rate, track how widely and effectively the model is being used. Finally, return on investment (ROI) calculations, considering cost savings, revenue generation, and time savings, evaluate the financial success of AI deployment. Together, these metrics provide a comprehensive view of how well an AI model performs and its value to the organization.

AI presents an unparalleled opportunity for pharmaceutical brand and sales leaders to drive business impact. By focusing on practical applications, addressing implementation challenges, and abiding by ethical principles, organizations will harness AI's potential to transform operations and deliver value to patients, physicians, and shareholders. Generative AI adds another layer of capability, enabling more creative and adaptive solutions. Moving beyond the hype, a strategic approach to AI adoption ensures sustainable success in an increasingly competitive landscape.

References

1. Strategy. Re-inventing pharma with artificial intelligence | Strategy& [Internet]. Strategy&. 2024. Available from: https://www.strategyand.pwc.com/de/en/industries/pharma-life-sciences/re-inventing-pharma-with-artificial-intelligence.html

About the Author

Vineet Rathi, Principal, Axtria Inc. – For over 20 years, Vineet has led large-scale, end-to-end commercial strategy consulting for the world's leading life sciences companies. Vineet brings a deep expertise in AI innovation to commercial process assessments. He has facilitated several large-scale market stakeholder workshops, thought leadership webinars, and conference presentations (including PMSA and WorldatWork) and has led successful industry benchmarking studies. Vineet's functional experience includes AI, omnichannel sales force effectiveness, alignment design, call planning, dynamic targeting, quota setting, and incentive compensation across top life sciences companies.

PMSA

PHARMACEUTICAL MANAGEMENT
SCIENCE ASSOCIATION

www.pmsa.org