



Patient Claims versus EHR Data – How to Choose Between the Two?

December 2016

Patient Claims versus EHR Data – How to Choose Between the Two?

Why is this data issue an important topic?

The movement toward measuring patient outcomes resulting from a specific drug therapy has caused researchers to look at different datasets than previously analyzed. Two patient-based datasets come to mind: 1) electronic health records (EHR), and 2) claims data. These two data sources will each be suited to answer different business questions. The topic this white paper will review is simply which dataset is more appropriate for answering certain types of business questions, look at some cases from the academic literature that apply each data, and implications for pharma commercial analytics.

Fundamental reasons are driving this push to look beyond traditional pharma data to sources like patient claims and EHR data:

1. The rise of performance-based managed care contracts is requiring pharma companies to demonstrate drug value in producing health outcomes and/or health economic benefits.¹ While performance based contracts are currently still relatively rare, even without these contracts, managed care plans are making decisions on formularies that are based on evidence of value. Managed care plans are also looking to guidelines and the treatment pathways being adopted by providers – which are also driven by evidence of outcomes and value.
2. A continued shift of pharma company drug portfolios to specialty medicines will place greater pressure on pricing, market access, and affordability, requiring new data sources to demonstrate value.² Given the proliferation of specialty therapies, there is a need for more nuanced understanding of how these medicines perform in a real world setting, such as drug-drug interactions, broader

information on side-effect profiles, performance in specific patient sub-populations, etc., that are not feasible to address in traditional patient studies.

3. The previous point echoes a call from the academic research literature that challenges the current pharma commercial model in needing to demonstrate value across the entire project/product life-cycle in light of changing product and market/environmental dynamics.³
4. This view has also been highlighted at recent PMSA (Pharmaceutical Management Science Association) annual meetings, such as a keynote address by then president and CEO of PhRMA in 2015.⁴ He noted the need for companies to demonstrate value by reducing gap between what it costs to bring about new drug innovation and society's willingness to pay for that innovation, adopt different ways of thinking and new science, and view future pharma issues in an interdisciplinary fashion.
5. The analysis of patient claims and EHR data was also recognized in a survey of pharma industry practitioners in sales force science (from both pharma companies and consulting organizations).⁵ Practitioners noted that emerging sales force strategy and big data developments must include analyzing health/economic outcomes, a different set of analytics needed to drive these outcomes, and on new patient-level data sources, such as patient claims and EHR data.

In short, the pharma industry is undergoing rapid changes, where as noted in previously published white papers in this series, analytics that demonstrate value to payers and patients will be at the center of critical decisions that determine drug success in the marketplace.



What is patient EHR and claims data?

EHR data is going to have the following features:

1. EHR will have the most comprehensive and complete information for the sample of patients it covers, e.g., all diagnostic tests, medical procedures, prescriptions over time, and likely includes some information on outcomes. However, questions have been raised regarding the consistency of methodological attempts to assess EHR data quality⁶ and caveats for use in comparative effectiveness research.⁷
2. A weakness of EHR data is the potential bias in the sample from which the data is derived, the source typically being one or a small number of health plans or health systems. For example, if a given EHR sample is sourced mainly from clinics, the EHR may underrepresent some of the difficult and advanced patient cases that

appear in academic centers. If biases are found, it may be possible to reweight and normalize the data to improve the reliability of estimates. While this reflects a limitation in EHR data, there are still numerous strengths given the comprehensive information for the patient sample that EHR data provides compared to claims data.

Given these features, EHR data should be superior for questions such as understanding of patient diagnosis and treatment, e.g., volume of diagnosed patients, treatment pathways for patients, progression between therapies, time on therapy, when different treatment regimens are used. EHR data should also allow for understanding differences in these treatment paradigms for different patient segments by age, gender, ethnicity, etc. Since EHR data should have information on physician specialties and movement between physicians, one can understand overall the involvement of

different physician specialties in delivery of care. All of this will be overall average behavior at a national level.

Claims data is going to have the following features:

1. Claims data will provide physician identifiable information, which will also allow for identification of the clinic or account, so that analysis can be done by account or any geographical area. This data will, however, have substantial gaps. For example, a view for a given patient will be incomplete because some claims will not be captured or patient drops out of data because of change in insurance plan, so one cannot get a completely accurate view of patient, and view of physician will be incomplete because only a portion of patients will be captured.
2. The percent capture reflects the set of patients that are typically seen having comprehensive information including both medical and Rx claims. This typically results in a 70% overall capture of Rx claims, 50% overall capture of medical claims, and thus an intersection of 20-30% of patients with both. The percent capture rate thus

requires careful analysis since the raw volumes are not accurate. Getting a good read on volumes can be done by triangulating with other available data sources.

Given these features, the claims data should be superior for questions related to targeting of accounts or physicians, relative opportunities at a geography level, and directional differences in behavior in different accounts and geographies.

Therefore, a broad and simplified summary is that the EHR data is better for the more strategic “where do I play” questions, e.g., how are patients currently treated, where do I want my product used, in what patients segments, which broad opportunities do I pursue, how do I position my product for each, etc. Claims data is better for the more tactical “how do I win” questions, e.g., who do I target, how well am I performing in each of these targets, what is the size of the opportunity in each of these targets, what resources do I allocate by target, etc. Finally, the current point-in-time assessments of claims versus EHR data sources (comparisons based mainly on the US market & data availability) are that the data landscape is very dynamic, and that EHR in particular are evolving in ways that will further



increase its value and applications over claims data given the following trends:

1. broader adoption of EHR in healthcare professional (HCP) offices.
2. standardization of information capture.
3. more of EHR data is becoming publicly available (as well as more robust patient samples) for use in research.
4. improvement in feature extraction techniques to derive valuable information from unstructured physician notes.

Some examples of applying EHR and claims data from the academic literature

This section is not meant to be an exhaustive review but show examples how each data source was used in different analyses. Some examples using EHR are as follows. The use of electronic coded EHR data from a random sample of Medicare patient charts was found to achieve a much higher identification of diabetes patients (97%) than manually-derived EHR from administrative claims data (75%).⁸ A 2016 study noted the importance of improving therapeutic effectiveness and safety through electronically recorded longitudinal big healthcare data.⁹ Large EHR databases can identify sufficiently numbers of subjects for future pharmacogenomics studies for very rare conditions when coupled with the introduction of newer ICD-9 codes.¹⁰ Lastly, and for balance, some research suggests that the ultimate vision for using EHR data for post-market drug surveillance has not yet been achieved.¹¹ The morale here is that care, understanding data construction, and potential biases need to be carefully weighed when conducting analysis using EHR data.

Some examples using claims data are as follows. Medicare claims data was used to look at national trend changes in health outcomes and medical services post-implementation of Part D.¹² Another study looked at medication adherence and persistence to different antihypertensive drug classes using claims data from a German statutory health insurance scheme that insures about 90% of the population (thus it overcame potential coverage gaps typical in claims data).¹³ A 2015 study analyzed respondents greater than 65 using various claims data to measure factors that would be

indicative of frailty and thus could hamper an estimation of drug effectiveness and safety.¹⁴ Another 2015 study used Medicare claims data to measure comparative effectiveness of different doses of influenza vaccines on elderly US residents.¹⁵ Lastly, a 2016 study estimated the prevalence of drug side effects, health care resource use, and costs in patients with asthma.¹⁶ Interestingly, the paper noted qualifications of the results due to limitations in using commercial claims data. Previously noted studies analyzing Medicare claims data are more encompassing of the elderly patient population and total utilization of health care resources than commercial claims data for its covered population. Thus claims data is subject to gaps of unmeasured confounding variables, casting doubt on estimating desired relationships, an issue noted in the earlier section. Lastly and interesting is a comment by the authors that a retrospective study using electronic medical records linked to health care claims would allow for a well-controlled analysis, bolstering previously noted differences between EHR and claims data.

Finally, the shift to specialty medicines, and especially drugs for rare diseases or highly targeted personalized medicines catering to a very small patient population will place greater weight on using EHR over commercial (non-Medicare) claims data for comparative effectiveness analyses. The fragmentation of and challenges in merging different commercial claims data will mean insufficient patient populations and gaps in measuring resource utilization and other key control variables will mean moving more toward analyzing EHR data.

Implications of EHR and claims data on the future of pharma commercial analytics

The move toward specialty medicines will increase the need for pharma companies to demonstrate measures of value that alter the focus and type of commercial analytics to support ever-increasing performance-based contracts and/or adoption decisions. Commercial activities can be broken down into seven buckets:

1. *Commercial Model Design* - the go-to-market approach and model design necessary to achieve all company strategic goals, but dependent on the drug technology

of the project/product portfolio that can be successfully developed and tactically executed in an efficient fashion to deliver optimal results while mitigating external threats and positioning the company to take advantage of opportunities (e.g., define metrics to determine success and how company resources will be positioned/organized/coordinated to ensure achievement of stated company goals). A significant change from previous and traditional commercial model design approaches will be ensuring success at the payer and patient levels at the center of biopharmaceutical go-to-market strategy, where all other analytics and data management activities playing support roles to these two areas.

2. *Payer Analytics* - focused on managed markets (e.g., private third party commercial and public drug plans), analyzing effects from changes in plan design, and their relationship to sales, marketing, and patient outcomes.
3. *Patient Analytics* - focused on analyses generated from real world evidence (RWE) and patient-level data (e.g., EHR and claims data) on outcomes (e.g., drug compliance and adherence, drug costs, treatment costs, health outcomes, cost-effectiveness) resulting from drug utilization.
4. *Sales Analytics* - focused on processes and outcomes related to ensuring optimal sales force investment efficiency and result effectiveness (e.g., sales force strategy outcomes, territory alignment, call planning, objective setting, incentive compensation, sales performance metrics, sales reporting).
5. *Marketing Analytics* - focused on processes and outcomes related to ensuring optimal brand performance throughout the entire lifecycle (e.g., emerging brand status, pre-launch preparations, launch, growth, maturity, and post-patent expiration).
6. *Commercial Analytics Innovation Center* - focused on basic research activities designed to generate new management/marketing science methods for solutions to address future commercial problems faced across the entire project/product lifecycle using experimentation, collaborations with academic researchers, and other activities to encourage innovation.

7. *Cloud Information Management* - focused on speed, agility, and scale in association with managing new data sources, elastic infrastructure, data quality & accuracy, and actionable insight in support of activities in all of the preceding commercial analytics buckets.

Traditionally, these commercial analytics buckets were seen and conducted as more distinct and separate activities. Today, and increasingly in the future, these commercial analytics buckets are rapidly becoming interdependent activities. Moreover, outcomes from (2) payer and (3) patient analytics, utilizing EHR and claims data, will become the principal emphasis and drivers of all other commercial decisions. Commercial analytics will be increasingly addressing some of the same questions that traditionally have been in the domain of medical affairs, in particular, providing evidence-based improvements in health outcomes and cost effectiveness (HEOR, health economics and outcomes research). Rather than being a conflict with traditional commercial analytics, this merging with HEOR modeling analytics is a better alignment of necessary activities. As the measures of commercial success become more focused on 'value' and 'cost effectiveness', there is better alignment between commercial and medical affairs that can better address payer and patient decision-making.

The construction of the right commercial model design and the conduct of all remaining analytics in other areas to support payer and patient outcomes, means solving problems in the near future using commercial analytics requiring greater alignment among these activities. In particular, what will be required is an open system framework of thinking in solving commercial problems, a data environment constructed to support all of these activities, and a leadership approach and innovative analytics culture necessary to cultivate and sustain a competition advantage. The empirical methods to analyze EHR and claims data are also fundamentally different that more closely resemble HEOR modeling than those employed in traditional sales and marketing analytics. Thus new skills will need to be acquired by commercial analytic functions to adapt to analyzing RWE models and merge those learnings into the remaining commercial operation work streams.

References

1. Edlin M. Performance-based pricing for pharmaceuticals. *Managed Healthcare Executive*, published online 30 September 2015, available at <http://managedhealthcareexecutive.modernmedicine.com/managed-healthcare-executive/news/performance-based-pricing-pharmaceuticals?page=0,0> (accessed 18 November 2016).
2. IMS Institute for Healthcare Informatics. Medicine use and spending in the U.S.: a review of 2015 and outlook to 2020. Parsippany, NJ: April 2016.
3. Singh J and Jayanti R. Closing the marketing strategy-tactics gap: an institutional theory analysis of pharmaceutical value chain. In: Ding M, Eliashberg J and Stremersch S. (eds.) *Innovation and marketing in the pharmaceutical industry: emerging practices, research, and policies*. New York: Springer, 2014, pp. 710-735.
4. Castellani J. Keynote presentation: The changing healthcare environment. Pharmaceutical Management Science Association Annual Conference 2015. Presentation on April 20, 2015, Arlington, VA.
5. Chressanthi G and Mantrala M. Pharmaceutical sales force science: current and emerging trends, and issues for future research. Presentation to the Pharmaceutical Management Science Association Board. Las Vegas, NV: PMSA Annual Meeting; 17 April 2016.
6. Weiskopf N and Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association* 2013; 20: 144-151.
7. Hersh W, Weiner M and Saltz J. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical Care* 2013; 51: S30-S37.
8. Tang P, Ralston M, Arrigotti M, et al. Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: implications for performance measures. *Journal of the American Medical Informatics Association* 2007; 14: 10-15.
9. Schneeweiss S. Improving therapeutic effectiveness and safety through big healthcare data. *Clinical Pharmacology & Therapeutics* 2016; 99: 262-265.
10. Davis R, Gallagher M, Asgari, et al. Identification of Stevens-Johnson syndrome and toxic epidermal necrolysis in electronic health record databases. *Pharmacoepidemiology and Drug Safety* 2015; 24: 684-692.
11. Moore T and Furberg C. Electronic health data for postmarket surveillance: a vision not realized. *Drug Safety* 38: 601-610.

- 
12. Briesacher B, Madden J, Zhang F, et al. Did Medicare Part D affect national trends in health outcomes or hospitalization?: a times-series analysis. *Annals of Internal Medicine* 2015; 162: 825-833.
 13. Schulz M, Krueger K, Schuessel K, et al. Medication adherence and persistence according to different antihypertensive drug classes: a retrospective cohort study of 255,500 patients. *International Journal of Cardiology* 2016; 220: 668-676.
 14. Faurot K, Funk M, Pate V, et al. Using claims data to predict dependency in activities of daily living as a proxy for frailty. *Pharmacoepidemiology and Drug Safety* 2015; 24: 59-66.
 15. Izurieta H, Thadani N, Shay D, et al. Comparative effectiveness of high-dose versus standard-dose influenza vaccines in US residents aged 65 years and older from 2012 to 2013 using Medicare data: a retrospective cohort analysis. *The Lancet Infectious Diseases* 2015; 15: 293-300.
 16. Luskin A, Antonova E, Broder M, et al. Health care resource use and costs associated with possible side effects of high oral corticosteroid use in asthma: a claims-based analysis. *ClinicoEconomics and Outcomes Research* 2016; 8: 641-648.



George A. Chressanthis, Ph.D.

Principal Scientist
Axtria Inc.

300 Connell Drive, Suite 5000
Berkeley Heights, NJ 07922

Email: george.chressanthis@axtria.com



Randy Risser

Principal
Axtria Inc.

300 Connell Drive, Suite 5000
Berkeley Heights, NJ 07922

Email: randy.risser@axtria.com

Contact Us

+1-877-9AXTRIA
info@axtria.com

Disclaimer

Axtria® understands the compliance requirements behind personalization and we do not work with any personally identifiable data that can identify an end-customer of a business.

We have the strictest data security guidelines in place as we work with businesses to improve the experience for their customers.



www.axtria.com



facebook.com/AxtriaInc/



info@axtria.com



linkedin.com/company/axtria



[@AxtriaConnect](https://twitter.com/AxtriaConnect)

Founded in 2009, Axtria® is a Big Data Analytics company which combines industry knowledge, analytics and technology to help clients make better data-driven decisions. Our data analytics and software platforms support sales, marketing, and risk management operations in the life sciences, finance, retail, and technology industries. We serve clients with a high-touch on-site and onshore presence, leveraged by a global delivery platform that focuses on reducing the total cost of ownership with efficient execution, innovation, and virtualization.

For more information, visit www.axtria.com

Follow Axtria on Twitter, Facebook and LinkedIn

Copyright © Axtria Inc. 2016. All Right Reserved