ARTICLE

Use of Open Claims vs Closed Claims in Health Outcomes Research

2023

Written By

COLUMBIA DATA ANALYTICS

KYTHERA

Executive Summary

SITUATION

In the ever-evolving landscape of healthcare, patient-centricity remains at the forefront of decision-making and quality assessments. Ensuring every patient's specific healthcare needs are met and their desired health outcomes achieved is paramount for life sciences companies. To achieve this, collecting data that guides patients along their healthcare journeys is crucial.

While randomized controlled trials (RCTs) have historically provided answers to many healthcare questions, the demand for clinical insights far outpaces the resources available for conducting RCTs. As a result, there is a growing interest in leveraging nonrandomized real-world evidence (RWE) as a supplement to RCT evidence, enhancing clinical decision-making.

RWE delves into the benefits and risks of treatments using real-world data (RWD), drawn from sources such as health insurance claims, electronic health records (EHRs), physician directories, and patient registries. These data sources are increasingly standardized and accessible for therapeutic research. RWE is gaining traction as a valuable tool for clinical, regulatory, and payer organizations, offering faster implementation compared to RCTs, especially in the context of rapidly spreading diseases.

> "Open claims … offer near-real-time insights into patient activities, with weekly updates."

OPPORTUNITY

In the area of outcomes research, closed-payer claims and open claims have emerged as key data sources to trace a patient's care journey, more holistically. Closed-payer claims, sourced directly from insurers or employer-collected health claims, have been traditional choices. Examples include Optum, MarketScan, IQVIA, and Premier for commercial claims and government datasets like Medicare and Medicaid. Closed-claims data capture nearly all events during a patient's enrollment period. Therefore, in all settings of care, closed payers' claims can provide a robust timeline of all events, which is critical for many types of outcomes research. Moreover, they generally contain a full record of all care that an insurance company has covered, which is often the complete care that a patient has received.

Closed-claims information has some drawbacks: First, most outcomes research looks for continuous care, which requires continuous enrollment of patients for the analysis. Although continuous enrollment is not typically a problem with government claims such as Medicare, it does decrease the sample sizes significantly for commercial claims. Another disadvantage of closed claims is the time it takes for this adjudication to occur; most closed-claims sources have a data lag time of three to six months. Pharmaceutical companies using commercial closed claims may have to wait several years until the sample size is large to analyze its recently approved medication.

Open claims data, although less common in research, has gained prominence. Open claims, acquired through practice management systems, clearinghouses, or pharmacy benefit managers, offer near-real-time insights into patient activities, with weekly updates. These datasets, potentially covering over 300 million patients, provide a wealth of variables for comprehensive statistical analysis. Claims information is typically received within days of a patient's medical or pharmacy encounter; as such, open claims provide a near-real-time view of patient activity. Most open-claims data are updated weekly and provide the most up-to-date information about the care.

Open claims include much of the same information content as closed claims—diagnosis codes, procedure codes, and drug identifiers. These datasets, relative to closed-claims data, are significantly larger (some can cover more than 300 million patients) and contain more variables for robust statistical analysis. The first challenge with open claims is patient-level completeness. As open claims are received from intermediaries between the provider of the medical service and the payer organization, not all intermediaries will serve all payers resulting in patient-level and claims-level data gaps or "missingness" and data duplication.

OBJECTIVE

Columbia Data Analytics developed a proprietary algorithm to remove duplications from open claims data and compared several outcome measures between the closed claims and open claims to see if missing a claim provides a significant issue for analysis. We hypothesized that since switching a clearinghouse for providers is rare, our estimates would be significantly close to the estimates for closed claims.

RESULTS

The results reveal that open claims data boast sample sizes 10 to 65 times larger than closed claims data. Moreover, these larger sample sizes yield comparable healthcare utilization percentages. This trend persists across various diseases, including dependence disorders, chronic conditions, childhood diseases, rare diseases, and oncological conditions.

“... open claims data boast sample sizes 10 to 65 times larger than closed claims data.”

CONCLUSION

Open claims data offer significant advantages for health outcomes research due to their larger sample size and real-time insights. While challenges exist, including potential data gaps, once cleaned and processed, open claims provide estimates akin to closed claims. This presents an opportunity for timely insights, particularly for new medications and rare diseases, in contrast to the substantial time lag associated with closed claims. Open claims are poised to reshape healthcare research, fostering innovation and timely decision-making in the field.

Use of Open Claims vs Closed Claims in Health Outcomes Research

Baser O[1,2], Baser E[2] , Yapar N[2] , Samayoa G[2] , Mete F[2]

[1]City University of New York, New York, NY, [2]Columbia Data Analytics, New York, NY

INTRODUCTION

An individual's specific healthcare needs and desired health outcomes are the driving forces behind all healthcare decisions and quality measurements. Collecting data that clearly guide patients' actions along their healthcare journeys is crucial for life sciences companies to provide patient-centered care.

Many questions concerning treatments and health care have been answered through randomized controlled trials (RCTs). However, the volume of clinical questions of interest greatly outweighs the available resources to conduct RCTs. Therefore, there is growing interest in whether nonrandomized real-world evidence (RWE) can be used to supplement RCT (Randomized Controlled Trial) evidence and aid in clinical decision-making.[1]

> "... the use of open-claims data, which are currently not commonly used in research, has also risen in prominence."

RWE focuses on the benefits and harms of treatments using real-world data (RWD), that is, routinely collected data relating to patient health status and/or the delivery of health care.[1] Sources of RWD, such as health insurance claims, electronic health records (EHRs), and patient registries, are becoming increasingly consolidated, standardized, and accessible for research on therapeutics. There is also growing interest in using RWE to support decision-making, both clinically and at regulatory and payer organizations.[1] RWE studies relying on existing data can often be implemented more rapidly than RCTs, providing an important advantage in the context of a new, rapidly spreading disease with high morbidity and mortality.

RWD can capture hospitalizations, causes of death, medication dispensing, and tests performed. Measurement of outpatient medication dispensing is typically well-captured through pharmacy claims.[1,2] An exception is out-of-hospital deaths, which are captured poorly in data sources not linked to administrative or other death records.[1] Also, while test results are also not typically recorded in claims, a positive result may be inferred from a test followed by an International Classification of Disease-10 (ICD-10) diagnosis code.

Outcomes research studies have been using closed-payer claims for RWE. Two kinds of data collection are available when providing RWE of a patient's care journey: closed-payer claims and open claims. These claims are provided directly by the insurer or a collection of employers sharing their employees' health claims with consulting services. Optum, MarketScan, IQVIA, and Premier are examples of commercial closed claims databases. Medicare, Medicaid, Veterans Affairs, and Department of Defense datasets are examples of government closed claims.

In the United States, the use of open-claims data, which are currently not commonly used in research, has also risen in prominence.[1,3] The administrative claims data typically used in pharmacoepidemiology are so-called closed claims (i.e., claims submitted to an insurance company, adjudicated, and paid)[1].

Closed-claims data capture nearly all events during a patient's enrollment period. Therefore, in all settings of care, closed payers' claims can provide a robust timeline of all events, which is critical for many types of outcomes research. Eligibility data show when the patient is in the system so that all care-related visits during that time frame are captured. Moreover, they generally contain a full record of all care that an insurance company has covered, which is often the complete care that a patient has received. Closed claims also include an enrollment file, which provides monthly information on whether a patient has insurance coverage, and, as such, sets an expectation for "observability," or which claims a patient incurs that a researcher would expect to see.

Despite its advantages, closed-claims information has some drawbacks:

First, most outcomes research looks for continuous care, which requires continuous enrollment of patients for the analysis. Although continuous enrollment is not typically a problem with government claims such as Medicare, it does decrease the sample sizes significantly for commercial claims. For example, the total sample size for MarketScan with 38,954,536 patients--would decrease to 16,773,183 million with a three-year continuous enrollment requirement. For Pharmetrics Plus Data, the sample size would decrease from

44,950,014 to 14,151,982 with three years of continuous enrollment. Since most of the studies require one-year identification periods, a one-year pre-index period, and a one-year post-index period, this could result in a significant decrease in diseases diagnosed or medications dispensed where there is a small number of patients to start with. On average, patients change insurers every 18 months (about 1 and a half years) and may disappear from the closed-claims files.

Another disadvantage of closed claims is the time it takes for this adjudication to occur; most closed-claims sources have a data lag time of three to six-months.[1] For example, Centers for Medicare and Medicaid Services (CMS) claims can lag from 6 to 9-months.[1] Medicare and Medicaid datasets from CMS have three- to four-year lag times. Commercial claims have a six-month lag.

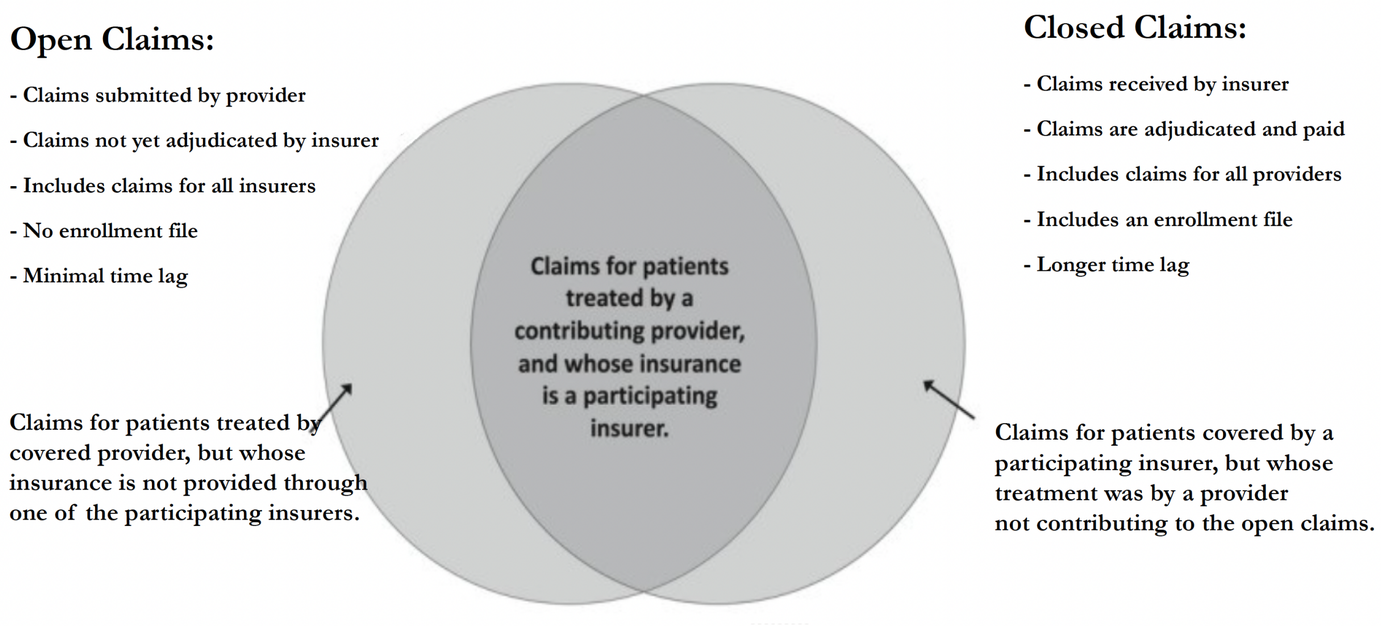
The Food and Drug Administration recently approved medications with post marketing commitments that show a serious risk related to the use of the drug or identify unexpected serious risks using real-world datasets. Therefore, a pharmaceutical company using commercial closed claims may have to wait several years until the sample size is large to analyze its recently approved medication. Most pharmaceutical companies are eager to show the advantage of their datasets to providers once their medication is on the market, and closed claims may delay these studies.

Open claims, on the other hand, are captured through practice management systems (the information systems that manage medical practices' scheduling, billing, and other internal functions), "switches" or "clearinghouses" (the companies that route claims from healthcare providers to U.S. insurers), or pharmacy benefit managers (companies that provide the link between pharmacies and insurance companies). These organizations often receive claims information within days of a patient's medical or pharmacy encounter; as such, open claims provide a near-real-time view of patient activity. Most open-claims data are updated weekly and provide the most up-to-date information about the care.

> Open claims include much of the same information content as closed claims – diagnosis codes, procedure codes and drug identifiers.

Open claims include much of the same information content as closed claims—diagnosis codes, procedure codes, and drug identifiers. When the care is provided at a specific healthcare setting, open claims capture the information and provide a glimpse into the patient's journey across several open-ended periods. These datasets, relative to closed-claims data, are significantly larger (some can cover more than 300 million patients) and contain more variables for robust statistical analysis.

However, it requires significant computing power and knowledge to process because the application of these claims to the outcomes research is new. However, open claims have several challenges. The first challenge with open claims is patient-level completeness. As open claims are received from intermediaries between the provider of the medical service and the payer organization, not all intermediaries will serve all payers, and there may be some selectivity as to which providers' data is included.[1] However, if patients tend to consistently seek care from the same providers, then this challenge should be mitigated.



A second issue is claims-level completeness, or the fraction of claims expected to be captured that are captured.[1] Claims-level completeness is time-dependent. Therefore, nearly all open pharmacy claims are observable within several days. Among medical and other claims that accrue over time, the majority are expected to be available within 21-days.

The enrollment file may also be problematic. One of the main disadvantages of open claims is the duplication of patients and missing claims since the enrollment file is null for the claims. Although closed claims provide an enrollment file that can be used to establish a denominator, open claims are more like EHR (Electronic Health Records) data in that they do not carry enrollment information. As such, the relevant denominator must be estimated through, for example, evidence of activity; that is, any patient on whom activity is observed within a specific time is included in the denominator.

To avoid biases due to under-inclusion of healthy individuals, definitions of "activity" that are more sensitive rather than more specific are recommended.[4] For a particular patient, the eligible person-time can be established similarly, where activity indicates that a patient's information would be expected to be captured, and lack of activity would exclude that patient's person-time. If these enrollment file issues are resolved, open claims would significantly contribute to health outcomes research as they provide large, more detailed, and current information than closed claims.

"... open claims data can provide information much earlier than closed claims which usually has a time lag of 6-8 months."

OBJECTIVES

We developed a proprietary algorithm to remove duplications from open claims data and compared several outcome measures between the closed claims and open claims to see if missing a claim provides a significant issue for analysis. We hypothesized that since switching a clearinghouse for providers is rare, our estimates would be significantly close to the estimates for closed claims.

METHODS

We used MarketScan and Pharmetrics Plus data as closed claims for 18 diseases using 2022 data as an identification period. After identifying each disease, we calculated the percentage of patients with hospitalization, emergency room, outpatient visits. Patients were continuously enrolled from January 2021 to December 2022. After applying our proprietary de-duplication technique, we identified the same group of patients from Kythera open-claims data, which are available continuously and updated weekly. The same outcome variables were created with the data. We used the chi-square test to compare the differences.

Table 1. Sample sizes and overall ratio from Open vs Closed Claims.

	Kythera					MarketScan					Pharmetrics Plus				
Diagnosis	n	Inpatient	Emergency	Outpatient	LOS	n	Inpatient	Emergency	Outpatient	LOS	n	Inpatient	Emergency	Outpatient	LOS
Alcohol dependence	613,823	19.28%	37.16%	98.04%	3.01	36,287	21.69%	35.71%	96.45%	3.87	14,174	16.76%	30.10%	95.58%	2.29
Ankylosing spondylitis	69,811	6.29%	16.50%	98.36%	0.81	7,875	6.03%	24.86%	99.06%	0.45	2,495	5.13%	16.95%	98.88%	0.40
Bladder cancer	142,627	12.59%	20.06%	98.98%	1.77	3,883	11.82%	25.83%	99.38%	0.97	2,201	13.86%	22.49%	99.05%	1.43
Cirrhosis	24,481	8.50%	16.68%	98.46%	1.17	1,509	9.34%	25.51%	99.54%	0.88	582	7.39%	19.07%	99.48%	1.37
Chronic kidney disease	2,447,280	19.16%	27.54%	99.21%	3.85	82,017	15.64%	27.43%	99.02%	2.03	35,926	14.45%	22.70%	98.72%	1.97
Crohn's disease	254,126	9.01%	19.39%	98.09%	1.22	26,586	10.28%	26.19%	98.65%	0.96	9,075	9.31%	19.79%	98.28%	0.88
Cystic fibrosis	22,704	9.30%	14.24%	96.47%	1.59	1,649	13.46%	24.01%	99.03%	2.08	525	11.05%	17.33%	98.29%	1.14
Endometriosis	187,539	5.32%	22.95%	98.56%	0.40	31,405	6.17%	30.43%	99.18%	0.36	8,461	4.72%	21.34%	98.83%	0.30
Hepatitis B	98,414	5.88%	12.53%	97.95%	0.94	8,175	4.24%	14.45%	95.11%	0.43	2,411	3.94%	9.00%	94.53%	0.52
Leiomyoma of uterus	460,347	5.55%	20.22%	98.75%	0.57	92,494	4.37%	24.58%	98.74%	0.30	21,672	3.74%	15.85%	98.33%	0.28
Lupus	51,794	8.46%	20.60%	98.75%	1.26	5,333	7.43%	27.66%	98.93%	0.72	1,500	7.47%	20.00%	98.87%	0.85
Multiple sclerosis	231,609	8.75%	17.16%	97.93%	2.63	19,881	7.08%	23.99%	98.80%	0.62	6,182	5.95%	17.26%	98.56%	0.58
Myositis	94,037	9.99%	21.12%	98.67%	1.66	16,338	6.33%	24.95%	98.07%	0.64	5,750	5.17%	16.28%	97.74%	0.60
Opioid dependence	785,702	13.59%	30.77%	98.17%	2.06	24,773	16.77%	34.63%	98.11%	3.14	9,190	13.88%	28.05%	97.64%	1.94
Rheumatoid arthritis	267,057	7.73%	17.34%	98.94%	1.02	25,622	6.05%	23.17%	99.55%	0.48	8,994	5.85%	16.47%	99.29%	0.45
Schizophrenia	334,399	25.48%	36.64%	97.90%	9.38	4,357	29.97%	39.84%	97.13%	6.93	1,246	26.48%	36.12%	96.31%	5.24
Type 1 diabetes mellitus	611,653	12.06%	20.88%	98.08%	2.39	53,499	10.42%	23.87%	98.48%	0.97	17,725	8.72%	17.77%	98.38%	0.83
Ulcerative colitis	274,323	9.55%	19.14%	97.88%	1.52	31,775	8.59%	23.93%	98.17%	0.83	10,504	7.60%	18.30%	97.83%	0.78

RESULTS

Table 1 presents the results. The sample size of open claims data was 10 to 65 times larger than closed claims data. 613,823 patients were identified with alcohol dependence in open claims data. In contrast, there were only 36,287 patients in the closed claims data in MarketScan and 14,174 patients in Pharmetrics Plus. Comparing health care utilization, we found comparable results; 19.28% (Kythera) of the patients were hospitalized with a diagnosis of alcohol abuse which was between the percentages found from the two closed claims data (16.76% and 21.69%).

When we compare health care utilization before diagnosis, we found comparable results. 19.28% of the alcohol abuse patients identified were hospitalized in open claims, in between the percentages found between the two closed claims (16.76% and 21.69%, Pharmetric Plus and MarketScan, respectively). 37.16% of the alcohol abuse patients identified were in the Emergency Room in open claims, which was between the percentages found between two closed claims (30.10 % and 35.71%, Pharmetrics Plus and MarketScan, respectively). 98.04% of the alcohol abuse patients identified were outpatients in open claims, which was between the percentages found between two closed claims (95.58% and 96.45%, Pharmetrics Plus and MarketScan, respectively).

Similar results were found in other dependence disorders, such as Opioid dependence. 785,702 patients were identified with Opioid dependence in open claims data, whereas there were 24,773 patients identified in MarketScan and 9,190 in Pharmetric Plus. There were similar results in the comparison between the claims regarding healthcare utilization. We only found that for outpatients' data, the difference was remarkably close (98.17%, 98.11%, and 97.64%, Kythera, MarketScan, and Pharmetrics Plus, respectively).

Similar trends were found in other chronic diseases such as Chronic Kidney Disease, Cirrhosis, Chron's Disease, Ulcerative Colitis, Lupus, Multiple Sclerosis, Myositis, and Rheumatoid Arthritis. The outpatient's data, the difference between them was remarkably close. Furthermore, in the other aspects of healthcare utilization such as inpatients and emergency rooms data the open claim percentage will always be between the closed claims percentage.

In childhood diseases such as Type 1 Diabetes Mellitus, we found a similar trend. 611,653 patients were identified with Kythera, whereas 53,499 patients were identified with MarketScan, and 17,725 were identified with Pharmetric Plus. 20.88% of identified patients were in the emergency room, between the percentages found in the two closed claims (17.77% and 23.87 %, Pharmetric Plus and MarketScan repecitvely).

In rare diseases such as Ankylosing spondylitis, psychiatric disorders such as Schizophrenia, gynecologic conditions such as Endometriosis or Leiomyoma of the uterus, and infectious diseases such as Hepatitis B, we found a similar trend with the total number identified and regarding health care utilization.

We even proceed to identify oncological diseases such as Bladder Cancer and found a similar trend. For Bladder Cancer patients, we identified 142,627 patients with Kythera (open claims), which was about 64 times larger than data found in closed claims (3,883 and 2,201, MarketScan and Pharmetric Plus, respectively). For healthcare utilization, we found a similar trend 12.59% of the Kythera Bladder cancer identified patients were inpatients, which was in between the percentages found in the two closed claims (13.86% and 11.82%, Pharmetric Plus and MarketScan, respectively).

CONCLUSION

Open claims with a bigger sample size and more current provide essential advantages for health outcomes research studies. With open claims, there is a chance of missing claims due to data structure, but once cleaned, it can provide estimates close to closed claims. Open claims data compared to closed claims data regarding healthcare utilization was found to be close to the range, regardless of the etiology of the disease. Therefore, especially for new medications and rare diseases, open claims data can provide information much earlier than closed claims which usually has a time lag of 6-8 months.

REFERENCES

1. Baser O, Samayoa G, Yapar N, Baser E, Mete F. Use of Open Claims vs Closed Claims in Health OutcomesResearch. JHEOR. 2023;10(2):44-52. doi:10.36469/001c.87538

2. Franklin JM, Lin KJ, Gatto NM, Rassen JA, Glynn RJ, Schneeweiss S. Real-World Evidence for Assessing Pharmaceutical Treatments in the Context of COVID-19. Clin Pharmacol Ther. Apr 2021;109(4):816-828.doi:10.1002/cpt.2185

3. Steiner JF, Prochazka AV. The assessment of refill compliance using pharmacy records: methods, validity, and applications. Journal of clinical epidemiology. 1997;50(1):105-116.

4. Lane JC, Weaver J, Kostka K, et al. Risk of hydroxychloroquine alone and in combination with azithromycin in the treatment of rheumatoid arthritis: a multinational, retrospective study. The Lancet Rheumatology.2020;2(11):e698-e711.

5. Rassen JA, Bartels DB, Schneeweiss S, Patrick AR, Murk W. Measuring prevalence and incidence